National Records of Scotland

Administrative Data Based Population Estimates Scotland

Statistical Research

Published on 14 December 2021

Disclaimer: These statistical research outputs are **not the OFFICIAL STATISTICS** for Population Estimates for Scotland. The Official Statistics can be found in the statistics and data section of National Records of Scotland's website.

This publication reports on the results of research into how population estimates might be produced using a range of administrative data.

Any presentation or use of these research outputs should make clear to users the nature and purpose of the statistics.

Contents

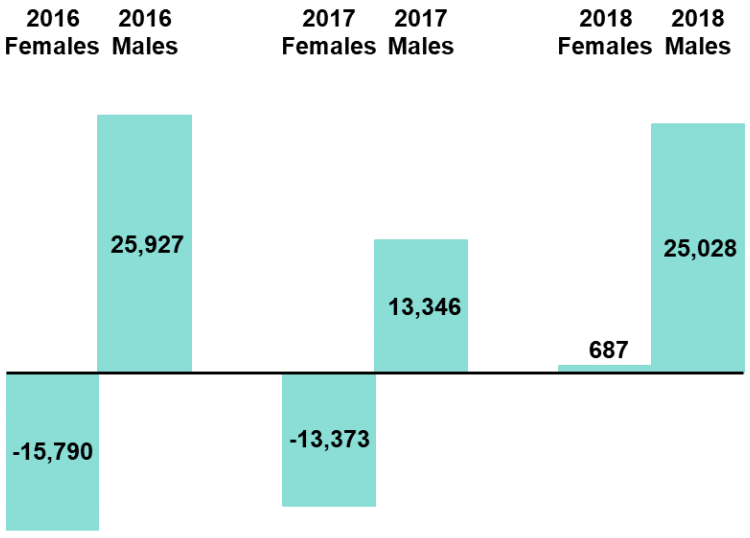
1. Key Findings

The aim of this statistical research is to create population estimates using only a range of administrative data. This research is part of the evidence base to support future developments in Scotland's demographic statistics.

This research is compared with the National Statistics Mid-Year Population Estimates (MYE), which start by using the latest census population estimates and then adjust for future years using data on births, deaths and migration. MYE are published by the National Records of Scotland each year.

- The Administrative Data Based Population Estimates (ABPE) for Scotland are higher than the Mid-Year Estimates (MYE) for 2016 and 2018, and very similar to the MYE for 2017.
- For all three years, the ABPE has more males than the MYE. In 2016 and 2018, the differences between the ABPE and MYE are larger for males than for females.

Figure 1: Difference in the population estimates between ABPE and MYE by year and sex.



- Different age ranges show different patterns, although these patterns are roughly consistent over the three years:
    - For young adults (aged around 18–25) the ABPE are generally higher than the MYE for females, and lower for males.
    - For those aged 30–65 there is little difference between the ABPE and MYE for females, but for males the ABPE is notably higher.
    - For those aged 67+ the ABPE for females and males are both lower than the MYE.
- The ABPE is generally higher than the MYE in the most-deprived areas, and lower in the least-deprived areas. There is a higher difference between ABPE and MYE for males in the most-deprived areas. This is consistent throughout council areas, and in each of the three years. Males in the age range of 30–59 show the largest percentage differences.
- The ABPE is generally higher than the MYE in urban areas, and lower in rural areas.

2. Introduction

This statistical research publication presents population estimates produced from the counts of linked administrative datasets to estimate the population of Scotland in 2016, 2017 and 2018. This publication supersedes any outputs in the initial statistical research, which was based on the 2016 data.

These estimates **should not** be considered as a replacement for the National Statistics Publication for population estimates. If you require population estimates for any purpose, such as resource allocation, planning of services such as education and health, or to incorporate into other statistics, please use the latest mid-year population estimates available on the NRS website. The figures in this publication should not be used for these purposes.

The purpose of this publication is to report the results of the research comparing ABPE figures with the existing population estimates. This research is therefore an important step forward in our understanding of how current administrative data might be used to provide key demographic statistics.

3. Acknowledgements

The process of producing this research on population estimates from administrative sources has involved a number of organisations and individuals. The National Records of Scotland (NRS) Admin Data team would like to thank our data suppliers. The suppliers, along with the datasets they provided, are given in Table 1. More information on the datasets can be found in Section 9.

Table 1: Suppliers of the data used in the analysis, and the datasets supplied.

Supplier	Dataset	Description
Electoral Registration Officers (EROs)	Electoral Register (ER)	People registered to vote in Scotland
Higher Education Statistics Authority (HESA)	HESA	Higher education students studying or domiciled in Scotland
National Records of Scotland (NRS)	NHS Central Register (NHSCR)	People who are or have been registered with a GP in Scotland, or whose birth was registered in Scotland
	Vital Events	Births, deaths, marriages and civil partnership registrations
Public Health Scotland (PHS)	Health Activity (HA)	Patients who have interacted with the NHS in the previous three years
Scottish Funding Council (SFC)	Further Education Statistics (FES)	Further education students studying in Scotland
The Scottish Government (SG)	Scottish Pupil Census (SPC)	People enrolled in state funded schools in Scotland

The NRS Admin Data team would also like to thank colleagues at the Scottish Government and eDRIS (part of Public Health Scotland (PHS)) for their ongoing support with this project. We would also like to thank all the stakeholders and peer groups, who have contributed their expertise and knowledge to support this work.

4. Background

Aims of the project

The aims of the project are:

- To help inform future recommendations for the census beyond 2022. This includes investigating administrative data collected by public bodies and services, which could be used to augment, complement or replace NRS' data collected by a traditional census. This statistical research is part of the administrative data based population and household estimates project, which will provide evidence for the future recommendations.
- To improve the coherence of our population and migration statistics across the UK — working in partnership with the Office for National Statistics (ONS) and the Northern Ireland Statistics and Research Agency (NISRA). This includes work as part of a cross Government Statistical Service programme to transform international migration statistics (one of the key components of population change). We are also collaborating on our respective programmes in NRS, ONS and NISRA to improve how we produce population statistics through greater use of administrative data.
- To support discussion with data suppliers and stakeholders on the application of this work and receive feedback on these population estimates to inform future developments.

Why Statistical Research rather than Official Statistics

For producers of official statistics, such as NRS, the term 'Statistical Research' is used to refer to research that is at an early stage of its development and would not meet the requirements for official or experimental statistics. By using this term, NRS is able to formally publish material that can support further discussion and development. This publication presents the second iteration of our methodology and is therefore still regarded as statistical research. As with our first publication on Administrative Data Based Population Estimates, NRS have provided a voluntary adoption statement to show how the principles of the Code of Practice for Statistics have been followed for this publication.

Mid-Year Population Estimate Methodology

In Section 6 the results are compared with respective Mid-Year Population Estimates (MYE). Figures from the Centenarian publication have also been used to provide a further breakdown of those aged 90 and over. A full description of the methodology used for the MYE is published on the NRS website but a summary is provided here.

The MYE uses the 2011 Census as a base and uses a standard demographic method called the cohort component method. The cohort component method can be summarised as follows:

- Take the previous mid-year resident population and age-on by one year.
- Then estimate the population change between 1 July and 30 June by:
    - adding births occurring during the year;
    - removing deaths occurring during the year;
    - allowing for migration to and from the area.

Adjustments are also made for changes in some population groups that are not captured by the internal or international migration estimates, in particular, members of the armed forces and prisoners.

5. Revisions

The 2016 data was used in a previous publication: Administrative Data Based Population Estimates, Scotland 2016. Those estimates were created using version 1 of our methodology. With statistical research there is an expectation that outputs will change due to improvements in data and methodology. The 2016 estimates have therefore been revised using improved methodology for this publication (version 2), allowing comparisons between the versions. Version 1 2016 ABPE (from the previous publication) should not be compared directly with the 2017 and 2018 estimates from this publication, as they are not calculated using the same methodology.

Methodology

As part of the rationale for publishing as statistical research, we expect further refinements in methodology to occur in future releases. This may be due to the addition of more datasets, improvements in data linkage techniques, changes in who is included (based on a set of rules), or through the introduction of statistical estimation techniques.

The Administrative Data Based Population Estimates v2 Scotland 2016–2018: Methodology Report provides full details of the methodology behind the estimates presented here.

In summary, the data linkage and de-identification processes remain unchanged from the previously published research. Linking variables were derived from four pieces of information: name, sex, date of birth and postcode. All personal data (name, date of birth and postcode, and all variables derived from them) is de-identified to make it impossible to recover the original information. This means we cannot identify individuals. The de-identified datasets are sent to the National Safe Haven, and then linked together using the de-identified linking variables. (More information on this process was published in the previous publication).

Once the de-identified records have been linked together the links are analysed. Records that appear to represent to same person are pulled together into a group. Each group is given a UPID (Unique Person Identifier). Some UPIDs will represent

persons who are not in the Scottish population on the reference date. Therefore, to make the list of UPIDs closer to the list of persons living in Scotland on the reference date business rules have to be applied. The business rules trim down the list of UPIDs, for example removing persons who appear on the death registrations dataset. Changes to the business rules changes the final list of UPIDs and hence the population estimates.

In this version those business rules have been amended slightly from version 1. In version 2, the rules **exclude UPIDs** if:

- they appear on the NHSCR with a flag to say that they have died or moved out of Scotland (or out of the UK)
- they appear on the electoral register as an overseas voter
- age, sex or (new for version 2) location data is missing

And include cases only if one of the two following conditions are met:

- they appear on the NHSCR with a Scottish posting, and
    - on at least one of the other datasets or
    - they have an armed forces posting (that is, even if they do not appear on any other dataset)
- or they appear on the birth registrations dataset with a birth in the past year (that is, they are aged below one and not on the NHSCR)

Additionally, inclusion is dependent on age when counting appearance on the Health Activity dataset. Evidence from the Health Activity dataset is only taken if the most-recent interaction with the NHS is within a particular time period with respect to the reference date for the estimates. The time period thresholds depend on the person's age:

- 1.5 years for those aged 50 years and above
- 3 years for those aged 22–32 years (inclusive)
- 2 years for all other ages (0–21 and 33–49 years)

Full details of the business rules and why they were modified for this research is available in the methodology report provided with this publication.

For each person an age, sex and (de-identified) postcode is assigned. The Health Activity dataset is prioritised over NHSCR where there is conflict between different datasets on these values for de-identified location, age and sex. If information is not available from the Health Activity dataset then it is taken from other datasets.

NRS use administrative datasets where sex or gender is self-declared. The ABPE use the term sex for reporting but we are aware that this variable may contain sex and/or gender from the underlying dataset. For most UPIDs the sex information will be derived from the Health Activity dataset, which is itself an amalgamation of various health datasets.

From the de-identified postcode a lookup table is used to assign the council area, Scottish Index of Multiple Deprivation (SIMD) decile[1] and Urban–Rural Classification[2].There has been a change to the urban–rural lookup. In version 1 a lookup from (de-identified) postcode to the 2013/14 urban–rural classification was used. For version 2 a lookup from data zones[3] to the 2016 urban–rural classification was used. This allows a more-direct comparison with the published MYE, which are also based on data zones.

At this stage some further cleaning of the data takes place on the de-identified data:

- Correction of some de-identified postcodes for Health Activity 2017 data. Some people had the same de-identified postcode in the Health Activity 2016 and 2018, and NHSCR data over the three years, but a different de-identified postcode on Health Activity 2017. For these people the Health Activity postcode in 2017 was changed to be consistent with the NHSCR de-identified postcode and the Health Activity data for the other years.
- Move prison populations to the data zone the prison is in.

[1] Based on 2016 SIMD deciles.
[2] 2016 version.
[3] Data zones are the key geography for the dissemination of small area statistics in Scotland and are widely used across the public and private sector. Composed of aggregates of Census Output Areas, data zones are large enough that statistics can be presented accurately without fear of disclosure and yet small enough that they can be used to represent communities. They are designed to have roughly standard populations of 500 to 1,000.

- Records that appear in data zones where the residential buildings are believed to have been demolished are placed in another location if the person appears elsewhere on any of the datasets, otherwise they are removed.

Version 1 for 2016 used HESA data for 2015/16 and 2016/17 academic years. In Version 2 only the HESA dataset for the preceding academic year was used (for example 2016/17 HESA for 2017 ABPE). For FES the preceding year was also used (apart from 2016 where 2016/17 data was used as 2015/16 data was not available).

Once completed, this constitutes Scotland's Integrated Demographic Dataset (SIDD). This statistical dataset represents the people in Scotland at a particular time, without knowing their name, date of birth or postcode. The SIDD is the dataset used to estimate the population. The population estimates (ABPE) are counts taken from the respective SIDD for each year.

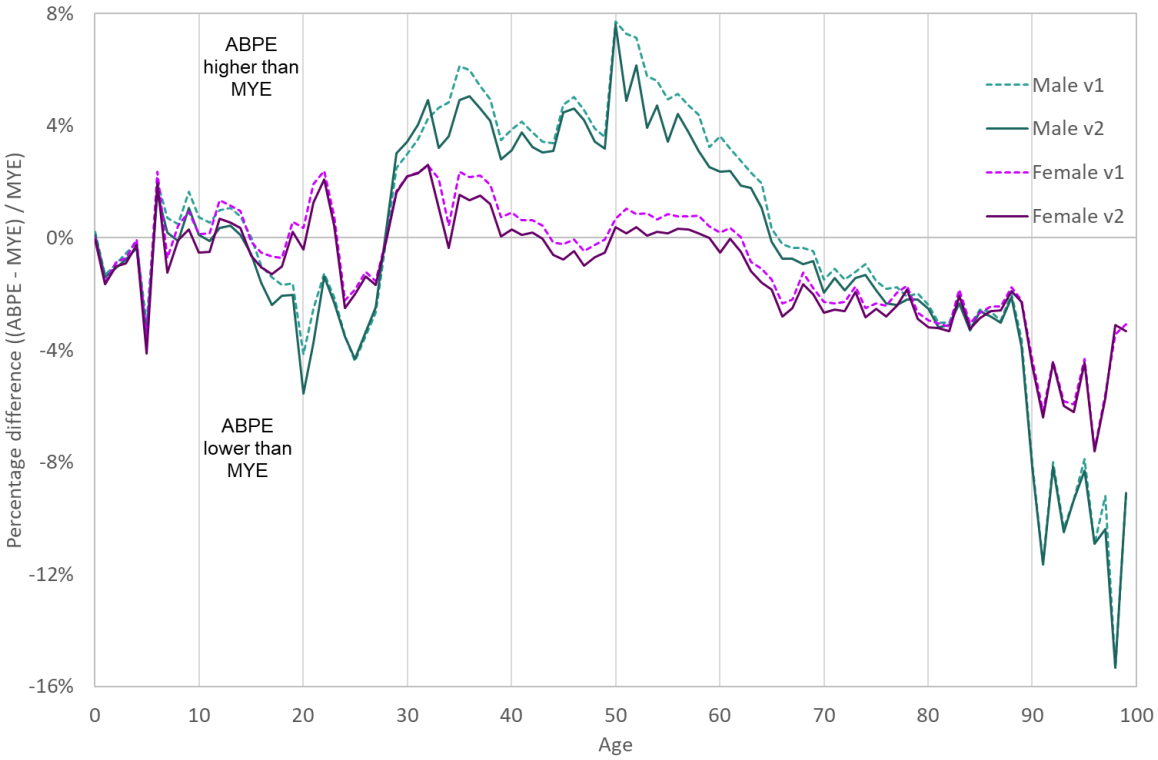
Comparison of Version 1 and Version 2 Estimates

Table 2 shows the overall difference between the ABPE and MYE for the two 2016 versions by sex. Figure 2 shows how the ABPE compares with the MYE for version 1 and version 2 for 2016 data for single year of age. Overall ABPE version 2 is nearer to the MYE for 2016. Further discussion on this comparison is included in Section 4.6 of the methodology report.

Table 2: Percentage difference between the ABPE and the MYE for version 1 and version 2 by sex, 2016.

	Females	Males	Total
Version 1	-0.1	1.5	0.7
Version 2	-0.6	1.0	0.2

Figure 2: Percentage difference between the ABPE and the MYE for 2016 by age and sex for version 1 and version 2 of the ABPE.



6. Analysis of Administrative Data Based Population Estimates, 2016 – 2018

Administrative data based population estimates (ABPE) outputs have been produced for:

- Scotland by single year of age and sex
- Council area by sex and single year of age and 5-year age bands
- SIMD deciles by sex and single year of age and 5-year age bands
- Urban-Rural classification by sex and single year of age and 5-year age bands
- SIMD deciles by 10-year age bands and sex and council area

These outputs have been compared with the associated National Statistics Publications covering 2016 to 2018:

- Mid-Year Population Estimates (MYE), 2016 to 2018
- Centenarians in Scotland, 2010 to 2020

Scotland Overall

The ABPE for each year can be seen to be broadly comparable with the MYE in 2016, 2017 and 2018, as shown in Table 3. Typically the population changes by tens of thousands each year. This is roughly similar in size to the difference between the ABPE and MYE. Even though 2017 has the smallest difference, we believe the ABPE for 2017 could be underestimated by around 10,200. Therefore the difference could be as high as +0.2% due to some known data issues in 2017 (more information on this is available in the council areas section).

Table 3: Difference between ABPE and MYE, Scotland 2016, 2017 and 2018.[4]

Year	ABPE	MYE	Difference
2016 (revised)	5,414,800	5,404,700	+0.2%
2017	5,424,800	5,424,800	0.0%
2018	5,463,800	5,438,100	+0.5%

[4] Numbers are rounded to the nearest hundred.

Scotland by single year of age

Figure 3: Comparison of ABPE-2018 and Mid-2018 Population Estimates for Scotland, by age.

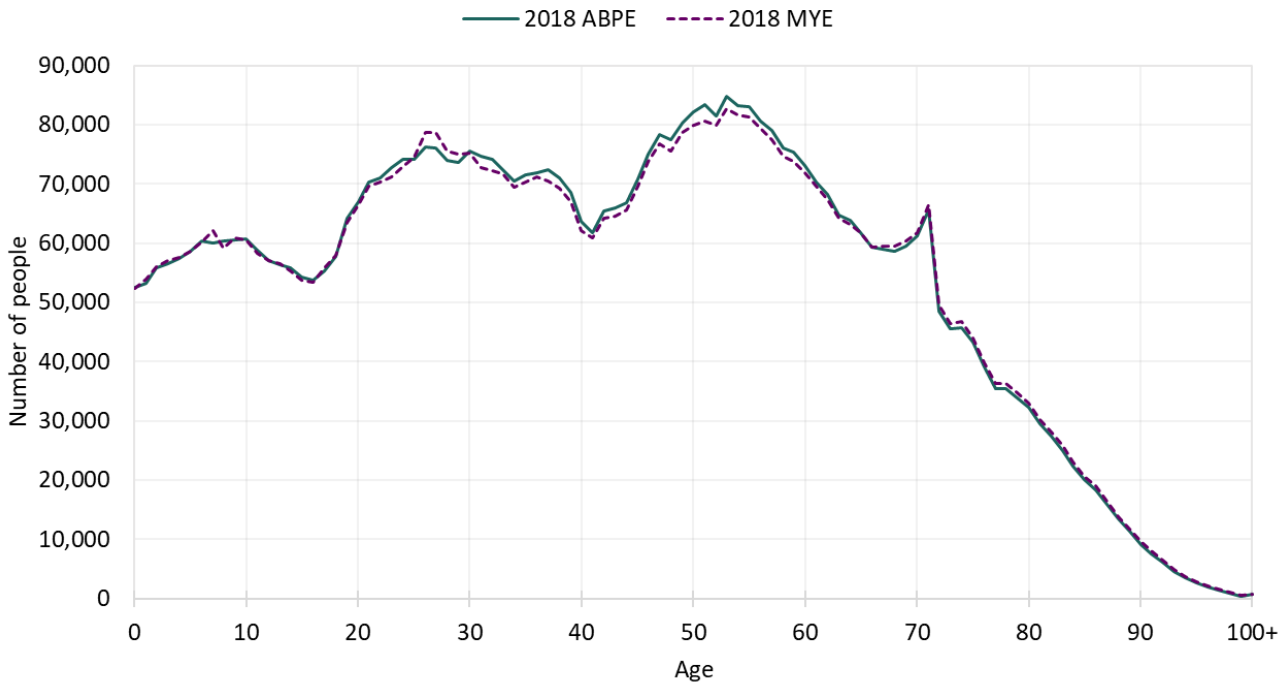


Figure 3, shows that the ABPE and MYE for 2018 are similar for all age groups. In 2018 the ABPE was 0.5% higher than the MYE. Figure 3 shows that it is mainly around ages 30–60 that the ABPE is higher than the MYE. Equivalent graphs for 2016 and 2017 (see charts) show similar patterns as 2018. This reflects the consistency in the quality of administrative data and demographic patterns in Scotland as a whole.

In Figure 4, the general demographic patterns of Scotland's population are replicated in each year. The peaks and troughs of the pattern generally phase forward as the relevant cohort ages. Around age 20, and above around age 75 are exceptions to this, where the pattern is very similar each year, rather than simply aging on. This suggests that around these ages the changes in the population are dominated less by people getting older, but by people entering or leaving the population at particular ages. This will plausibly be due to deaths above age 75, and immigration around age 20.

Figure 4: Comparison of ABPE for Scotland, 2016 to 2018 by age.

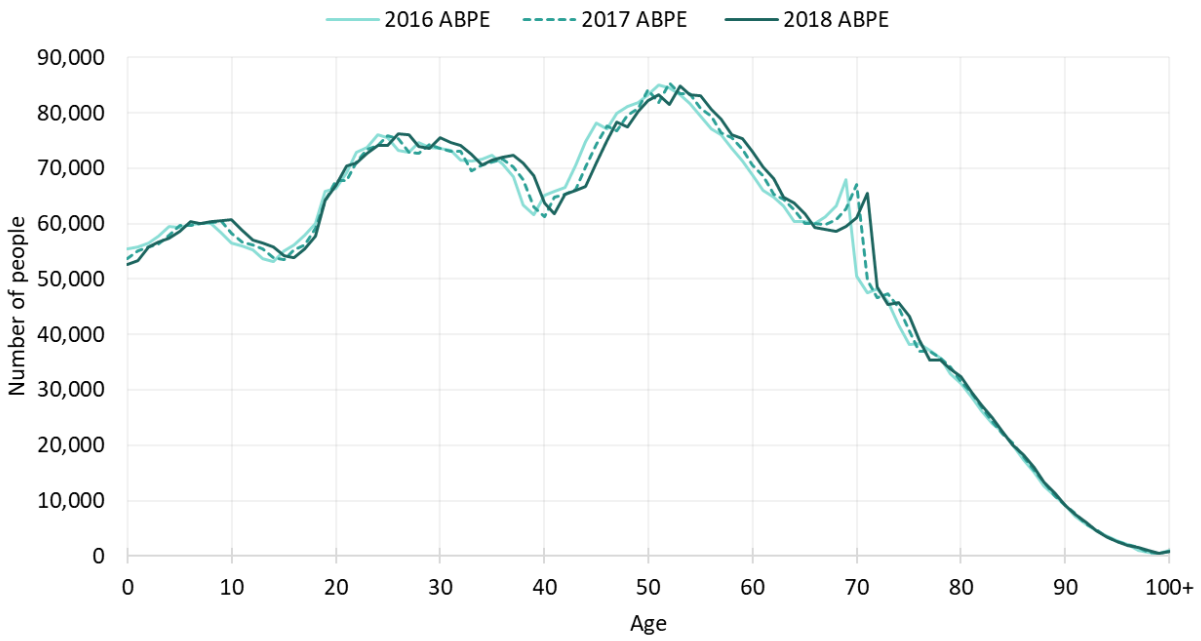
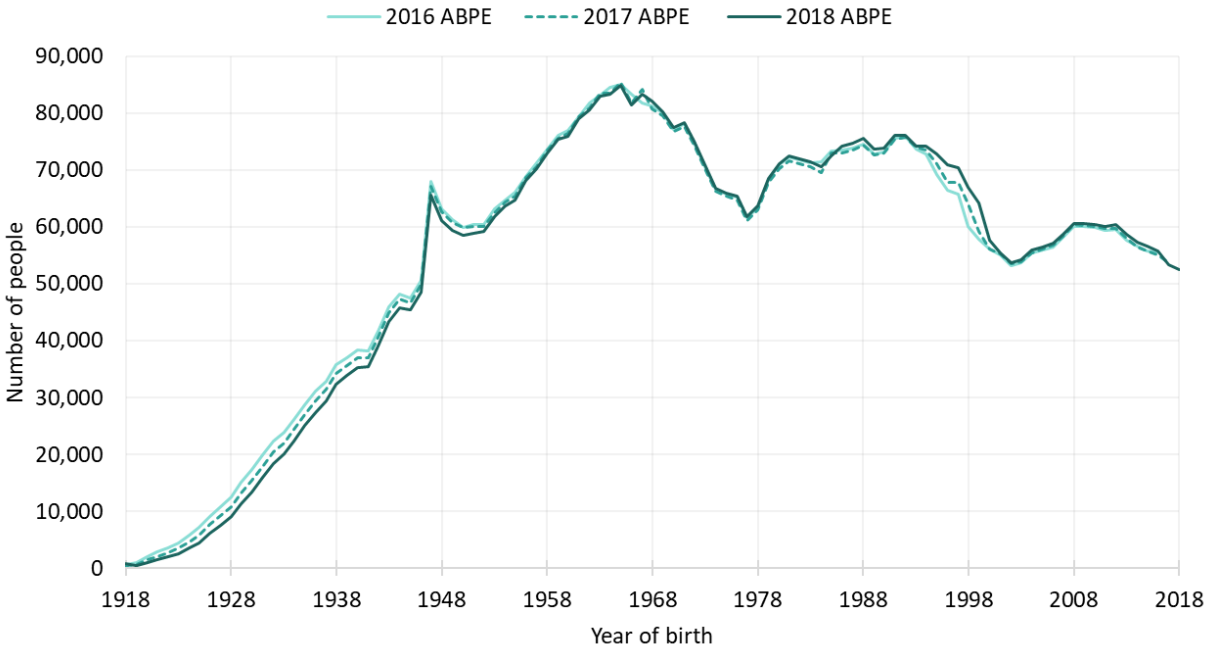


Figure 5: Comparison of ABPE for Scotland, 2016 to 2018 by year of birth.



The ABPE for Scotland for 2016 to 2018 is shown in Figure 5 by year of birth. This therefore shows how many people of a particular cohort appear on the ABPE each year. For most cohorts aged up to around 60 (born back to around 1958) the ABPE

estimates about the same number of people for each cohort each year. Cohorts aged around 20 (those born around 1998) increase in population each year. This shows that the ABPE is detecting people (including students) moving into Scotland and adding to the Scottish population for these cohorts. Conversely, estimates for cohorts aged above around 60 (born before around 1958) increasingly drop year on year, showing that the ABPE is detecting the natural change as people die.

Figures 4 and 5 together show that the ABPE for 2016, 2017 and 2018 are very similar either for particular ages or particular cohorts. It therefore appears that the ABPE are responsive to not just aging of cohorts, but changes within these cohorts due to migration and natural change.

Scotland by age and sex

Table 4 compares the ABPE and MYE by sex. This shows that the ABPE have more males than the MYE over the three years, and generally fewer females. These differences are explored in more detail.

Table 4: Difference between ABPE and MYE by sex, Scotland, 2016, 2017 and 2018.

	Female			Male		
Year	ABPE	MYE	Difference	ABPE	MYE	Difference
2016	2,761,407	2,777,197	-0.6%	2,653,433	2,627,503	1.0%
2017	2,771,130	2,784,500	-0.5%	2,653,648	2,640,300	0.5%
2018	2,790,039	2,789,349	0.0%	2,673,779	2,648,751	0.9%

Figures 6 and 7 show the percentage difference between the three years of ABPE by age and sex against the MYE. A notable difference is in the 30 to 65 age range, where the percentage difference is relatively small for females but the ABPE is consistently higher than the MYE for males. There are also divergent patterns to be examined in the 18 to 30 age range.

Figure 6: Percentage difference between ABPE and MYE by age for females.

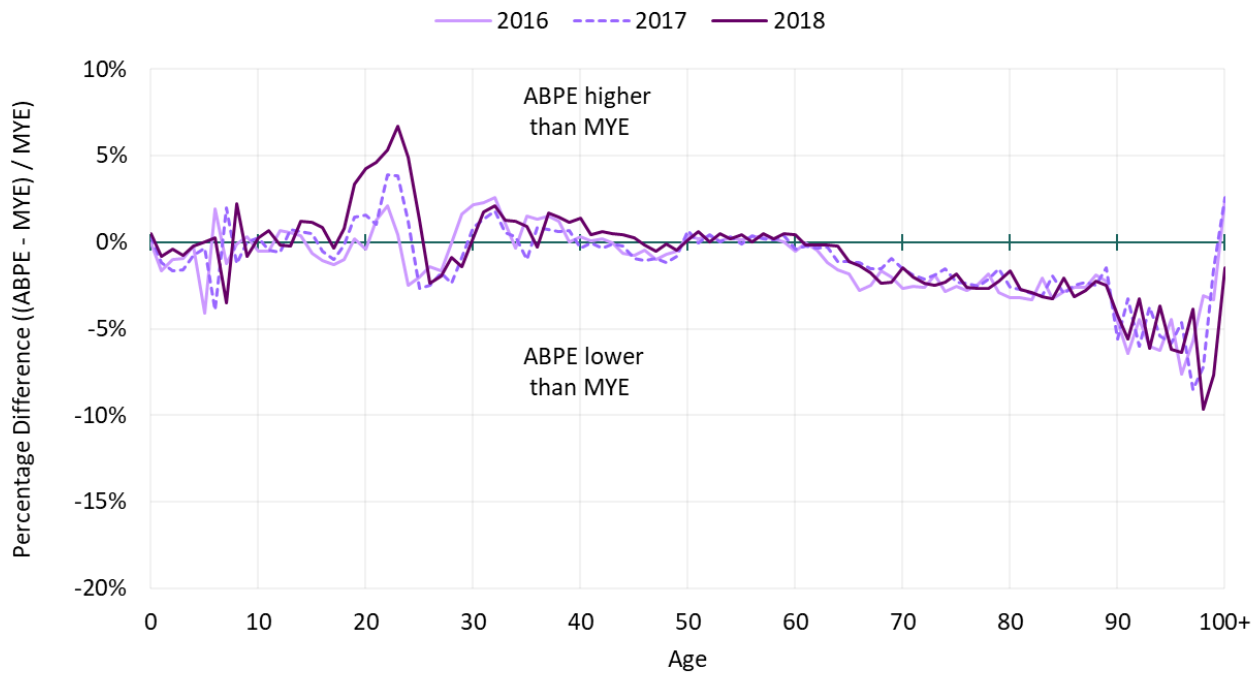


Figure 7: Percentage difference between ABPE and MYE by age for males.

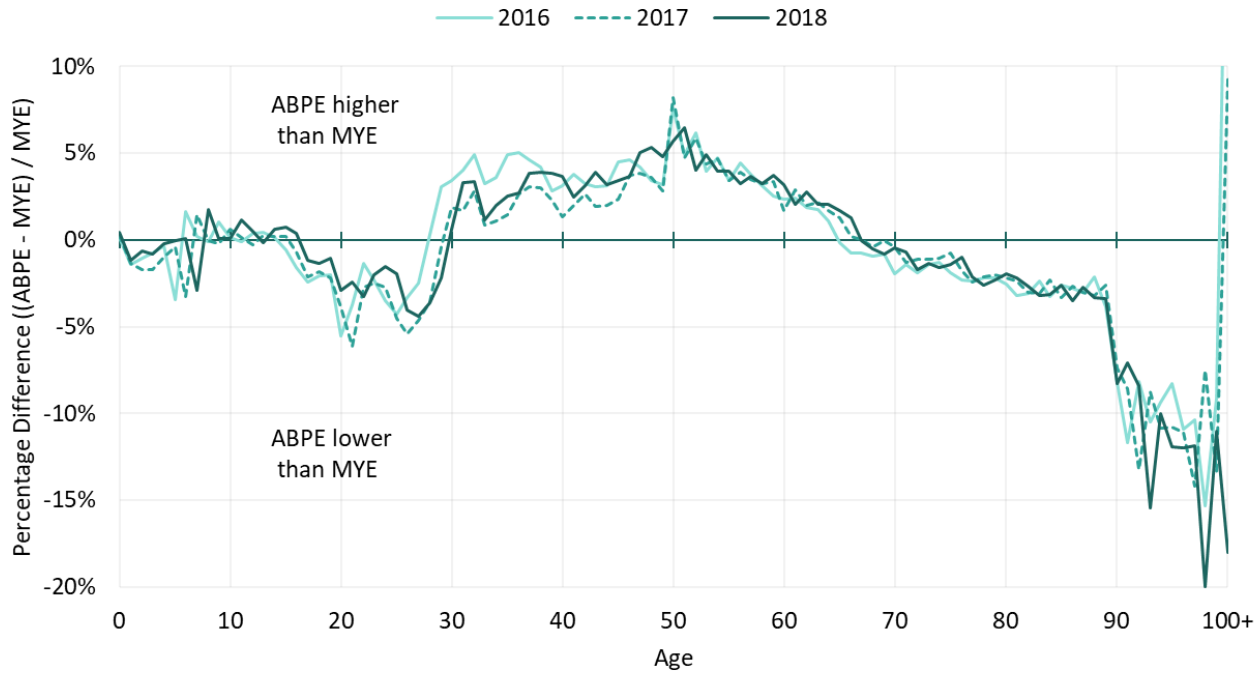


Figure 6, shows that the number of females over the 3 years are nearer to the MYE compared with males. For females between 30 and 65 years old, the biggest

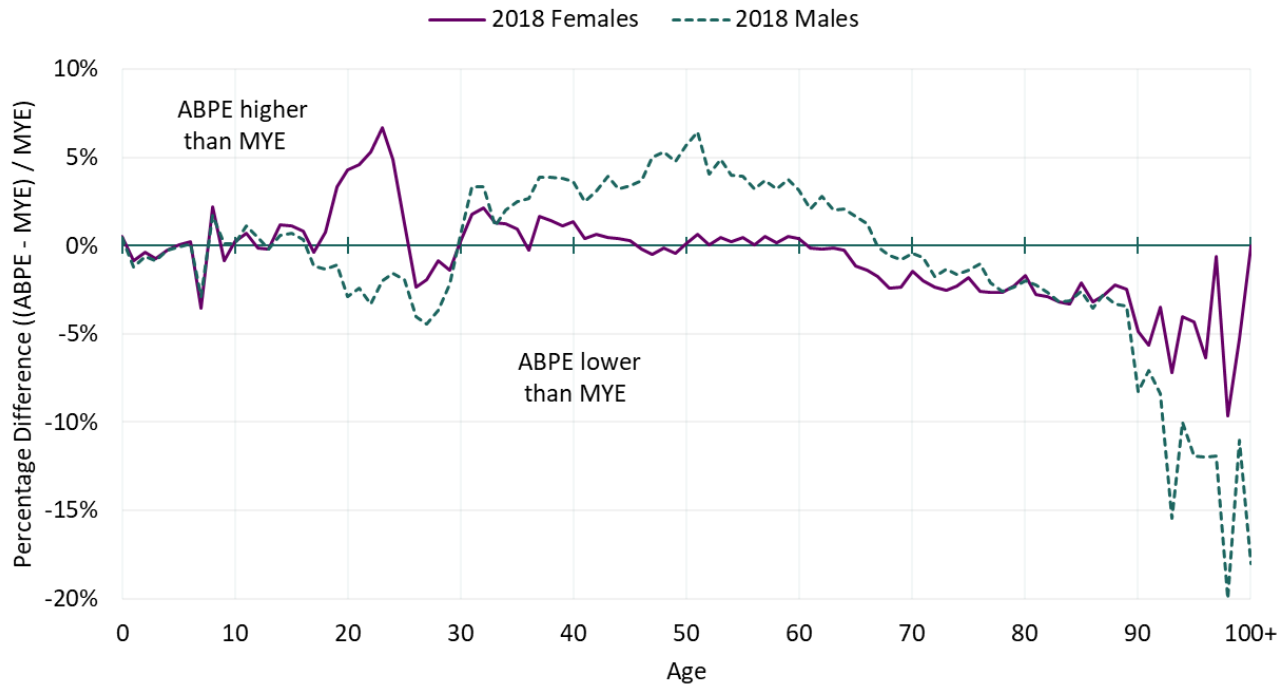
difference over the three years was at age 32 in 2016: 910 (2.6%) more females than the MYE. For males, between 30 and 65 years old, (Figure 7), the biggest difference was at age 50 in 2017, with 3,236 (8.2%) more males than the MYE.

The peak at age 50 (more pronounced for males, but also seen for females) is believed to be associated with bowel screening, which starts at that age. The biennial nature of that screening programme can be picked up in the ABPE patterns with mini peaks every two years beyond age 50, indicating that this is highly likely to be a contributing factor. This suggests that every two years the majority of Scotland's population are interacting with the NHS and this is reflected in the ABPE.

Figure 6, also shows that the ABPE has more females than the MYE around ages 19 to 23. This is not seen in the same age range for males as there tends to be comparatively fewer on the ABPE.

Although the percentage difference between the ABPE and MYE is very large for those aged 90 and above, there are very few people in these age groups, so this represents a small difference between the two estimates. For example, the 20% difference between males aged 98 is a difference of 48 people.

Figure 8: Percentage difference between ABPE and MYE by age and sex, 2018.



Females and males are shown separately in Figure 8 for 2018 to highlight the differences discussed above. Females tend to have more-recent interactions with the health service, and hence may be more likely to appear on the Health Activity dataset. Similarly, it may be that females are more likely to register with a Scottish GP when coming to live temporarily in Scotland, for example as a student, and hence may be more likely to appear on the NHSCR. It may also be that females are more likely to change their GP when moving into a new area, as females tend to be covered by more national health programmes earlier:

- Cervical screening is routinely offered to anyone with a cervix in Scotland between the ages of 25 and 64 every 3 years (during the period covered by this publication).
- Screening tests are offered to all pregnant women
- Contraception prescribed by healthcare professionals

Figure 9: Difference in population changes between 2016 to 2018 by cohort and sex, Scotland.

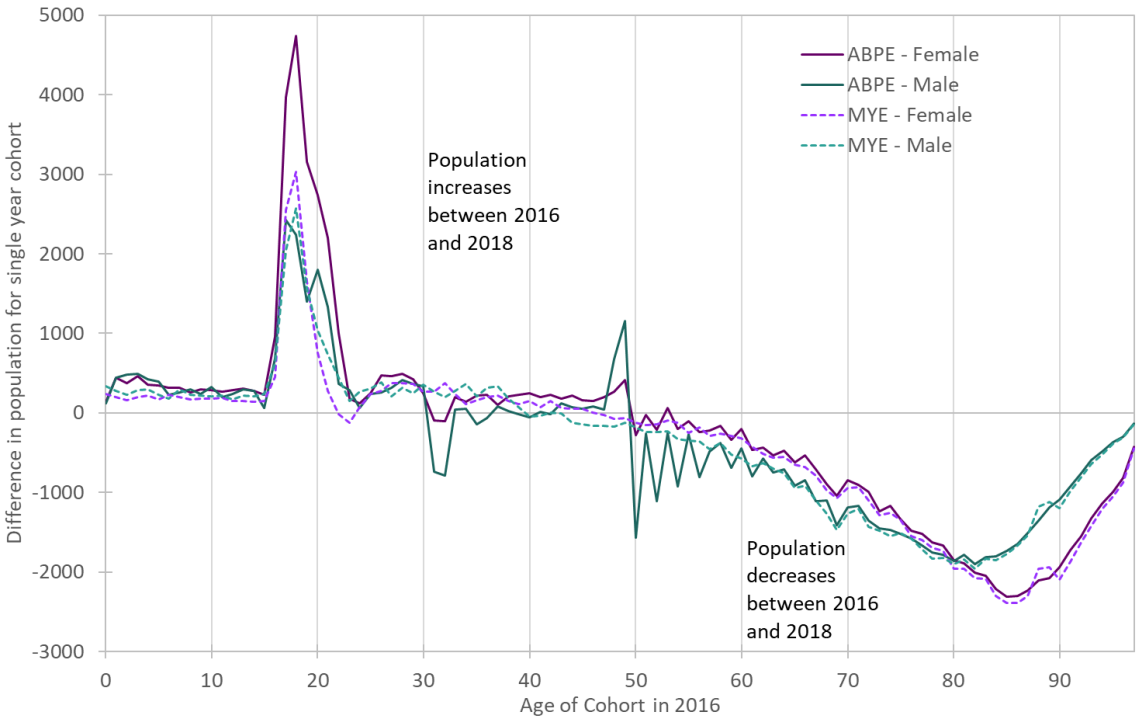


Figure 9 is based on cohorts by sex. The cohorts are taken at 2016 and age over time. For example the group of people aged 18 in 2016 will be the same cohort as

the group of people aged 20 in 2018. The size of each cohort in 2016 is then compared with the equivalent in 2018. By following a cohort the changes will not be due to ageing on, but rather by people entering or leaving the population.

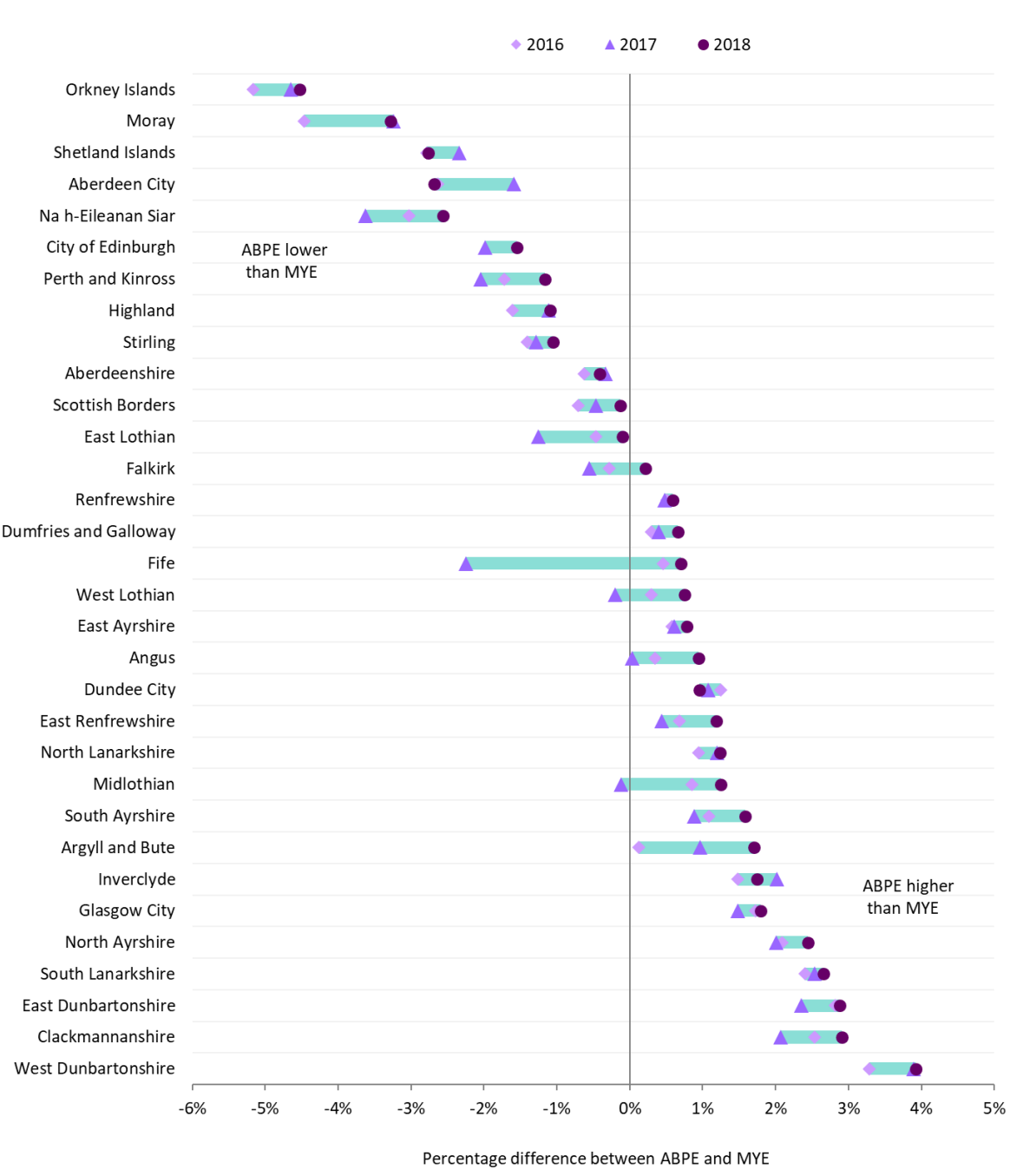
This shows that over the two years there are estimates for, the ABPE not only closely follows the MYE, but also follows the changes in the population by age and sex. Many of the sources of information that record changes in the population are the same for MYE and ABPE, for example births and deaths. NHSCR is used for internal migration in MYE and is a main part of the SIDD. We would therefore expect that the changes in the estimated size of each cohort to be similar when measured by the MYE as by the ABPE. This is indeed what is seen for most cohorts in Figure 9.

However, there are some notable differences. For males the most striking difference is for those around age 50. This will be due to the fluctuations in health activity due to the bowel screening programme. Both data sources show growth for cohorts aged around 20, but for females there is a substantially larger increase in the ABPE. This may reflect differences in how changes in student populations are estimated between the two methodologies.

Council areas

The comparisons between the ABPE and the MYE, broken down by council area are shown in Figure 10.

Figure 10: Percentage difference between ABPE and MYE by council area, 2016–2018, ordered by 2016 difference.



Three key points can be drawn from Figure 10. The first relates to how similar the ABPE and MYE are for the council areas. There are 11 council areas with the ABPE lower than MYE and 16 councils have an ABPE higher than their corresponding MYE. This was consistent over the three years. The percentage difference between the ABPE and the MYE for council areas range between -5.2% and 3.3%, -4.6% and 3.9%, and -4.5% and 3.9%, for 2016, 2017 and 2018 respectively. The ABPE is within ±1.0% of the MYE for eight of the council areas for all three years and 20 council areas are within ±2.0%.

The second point relates to how similar the difference between the ABPE and MYE is over the three years for each council area. The bar connects the minimum and maximum percentage differences for the three years. This shows the range of the percentage difference between ABPE and MYE. The smaller the range (the smaller the bar), the more consistency in the percentage difference over the three years. Council areas tend to have a smaller bar that connects the three years (the range), than their distance from the origin (the typical difference from the MYE). This suggests a degree of stability in the pattern of comparisons between the ABPE and MYE across the council areas.

Over the three years 25 council areas have a range smaller than 1 percentage point. However, while a small range indicates consistency in the difference between the ABPE and MYE, this (consistent) difference may still be substantial, in many cases more than 1%. For example, Glasgow City has differences of 1.7%, 1.5% and 1.8% for 2016, 2017 and 2018 respectively, a range of 0.3 percentage points. So while the difference is about the same each year, the ABPE is consistently higher than the MYE. Due care is required when interpreting the overall percentage differences.

The third point concerns the comparison for Fife in 2017 being notably lower than for 2016 and 2018 as shown in Figure 10. Most people on the SIDD are included because they satisfy the business rule requiring them to appear on the NHSCR and at least one other dataset. For adults the main datasets are Heath Activity and Electoral Register (ER). In 2017 Electoral Register data for Fife was not included in the analysis for 2017 due to a clerical issue (see the QAAD report for further details). People who appear on the NHSCR and ER but no other datasets would normally be

included on the SIDD. However, because the ER Fife was not linked to the other datasets any such people in Fife would not appear on the SIDD. This results in the ABPE for Fife being 2.2% lower than the MYE in 2017, compared with 0.5% and 0.7% higher in 2016 and 2018 respectively. Had Fife shown a similar comparison between the ABPE and MYE in 2017 as it did in 2016 and 2018 its ABPE estimate would have been around 10,200 higher.

Many adults will appear on NHSCR, ER and Health Activity. As people only need to appear on one of ER or Health Activity to be included, the methodology can largely cope with the loss of data from either ER or Health Activity. This meant that, although ER is a substantial dataset, the effect of it being absent for Fife was comparatively minor.

The 2017 issues were exacerbated by the issues with the 2017 Health Activity de-identified postcodes. That could result in links between the NHSCR and Health Activity potentially being missed. If that happened then it would appear that the person only appeared on the NHSCR, and so would not be included on the SIDD.

This has been an invaluable insight in the research showing how important the underlying quality of the data is. The quality of administrative data is an area that we are continually trying to improve, working alongside our data suppliers. It is considered a very important aspect of this research.

Figure 11: Distribution[5] of council area differences between ABPE and MYE by sex.

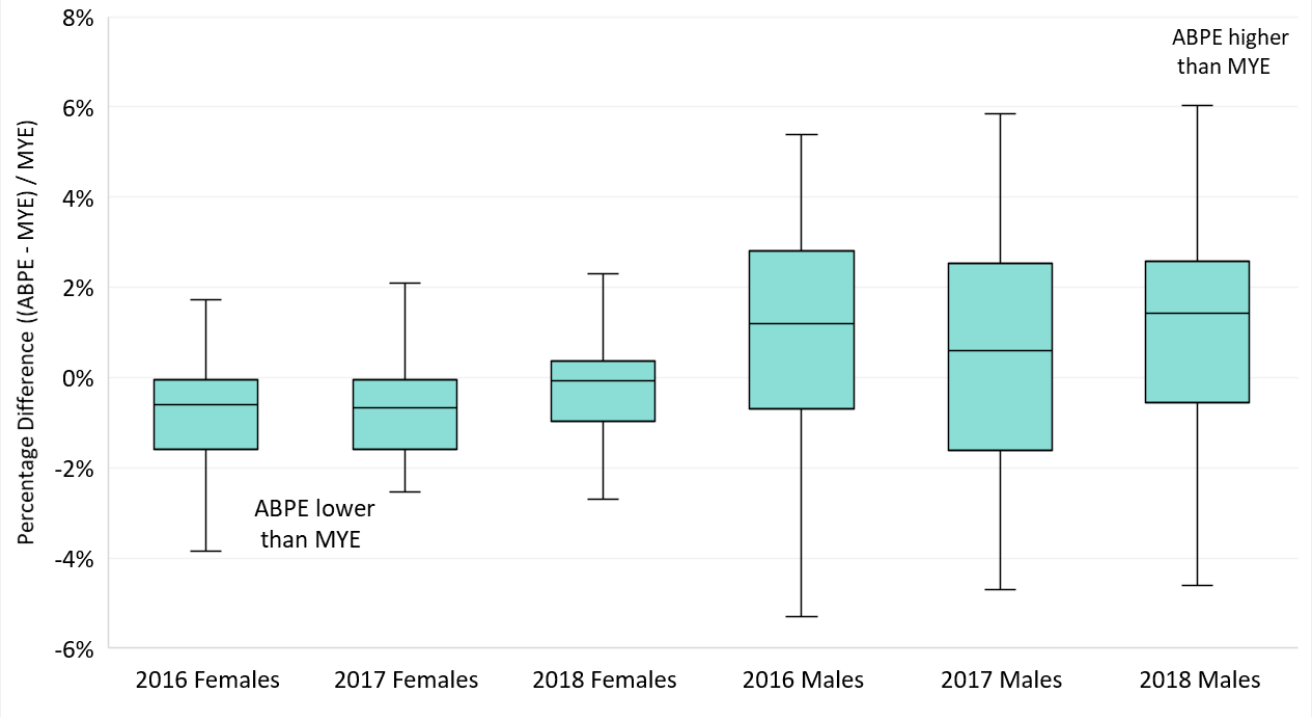


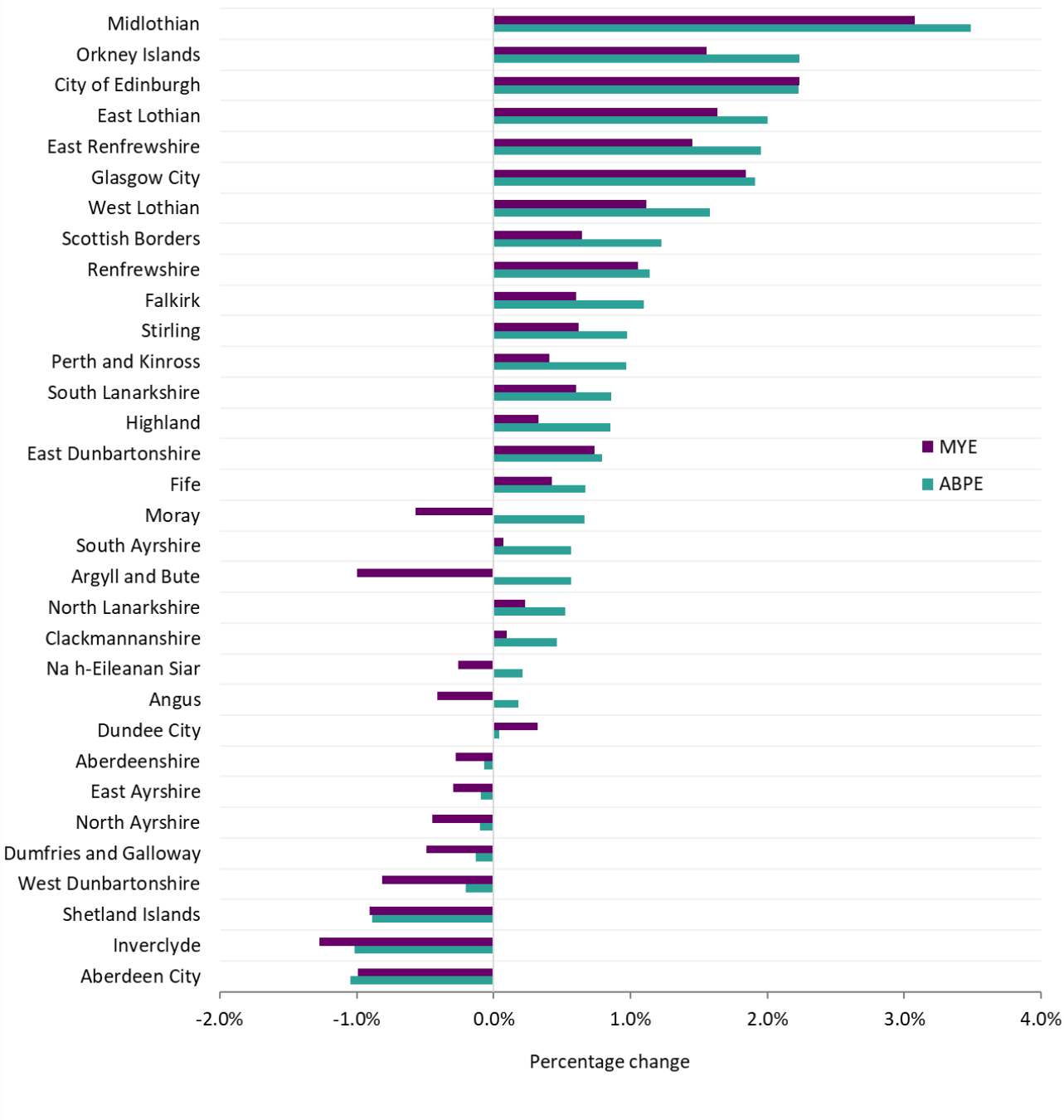
Figure 11 shows the distribution of the differences between the ABPE and MYE across council areas for females and males. As discussed previously, the ABPE tended to be higher than the MYE for males, while the comparisons were closer for females. This effect can also be seen in Figure 11, with males in council areas generally having the ABPE higher than the MYE. The median (shown by the central line within the boxes) for males is greater than 0% in all years, while females tend to have a lower median.

In addition to reflecting the overall difference between comparisons for males and females, Figure 11 shows how the variation in the comparisons between ABPE and MYE across council areas differs for males and females. Males consistently have greater variability in the percentage differences than females (as shown by the length of the box and whiskers in Figure 11). This means that for males some council

[5] The box shows the interquartile range, with the central line showing the median. The whisker shows the full range, excluding outliers. Outliers are defined as council areas that are more than 1.5 times the interquartile range below Q1 or above Q3, and are not shown.

areas will have the ABPE higher than the MYE, while other areas have the ABPE lower. For females the council areas have less variability in the difference between the ABPE and MYE. For females and for males the patterns in the differences between the ABPE and the MYE are broadly consistent across the years.

Figure 12: Percentage change for ABPE and MYE by council area, 2016–2018.



In addition to the differences between the ABPE and MYE in each year, the changes from 2016 to 2018 in each council area can be explored and compared between the ABPE and MYE. Figure 12 shows change in the population between 2016 and 2018 for each council area. The changes in the council areas for the four largest cities[6] are very similar for ABPE and MYE, or show the MYE increasing more than the ABPE. However, most other council areas that have a positive increase in MYE population between the 2 years tend to show a higher increase in the ABPE. Of the 12 council areas where there has been a decrease in population for the MYE, the ABPE decrease is smaller for eight council areas, and the other four have a positive ABPE increase.

The two council areas showing the largest difference between the ABPE and the MYE are Argyll and Bute, and Moray, which are two of the four council areas where the changes are in the opposite direction. In Argyll and Bute the ABPE increased by 492 people, while the MYE decreased by 480. The equivalent for Moray was +611 and -550. For both these council areas some, but not all, of the difference in the changes can be accounted for by the data zones containing MoD bases.

It is also possible to explore the change in population by sex and cohort between 2016 and 2018 for each council area, similar to Figure 9 for Scotland. In figures 13 and 14, Glasgow City and North Lanarkshire have been used as examples. As these cover a smaller population, there is more fluctuation than for Scotland. However, the changes in the ABPE still tend to follow the MYE pattern.

In Glasgow City, (Figure 13) there is a large increase in population for the cohort aged around 20. Many of the people who contribute to this increase are likely to be those who move to the city to study. For males the ABPE and MYE changes are broadly similar. The increases for females are larger than for males for both MYE and ABPE, but the ABPE shows larger increases than the MYE. This difference may be due to differences in the methodology for students between the MYE and ABPE, resulting in differences in the comparisons between ABPE and MYE for males and females for this age group, especially in 2018. It is known that the number of people

[6] The largest cities in Scotland are Glasgow, Edinburgh, Aberdeen and Dundee.

starting first degrees in Scotland increased between 2015/16 and 2017/18 by more for females (increasing by 1,890) than for males (increasing by 825)[7].

Figure 13: Difference in population changes between 2016 to 2018 by cohort and sex, Glasgow City.

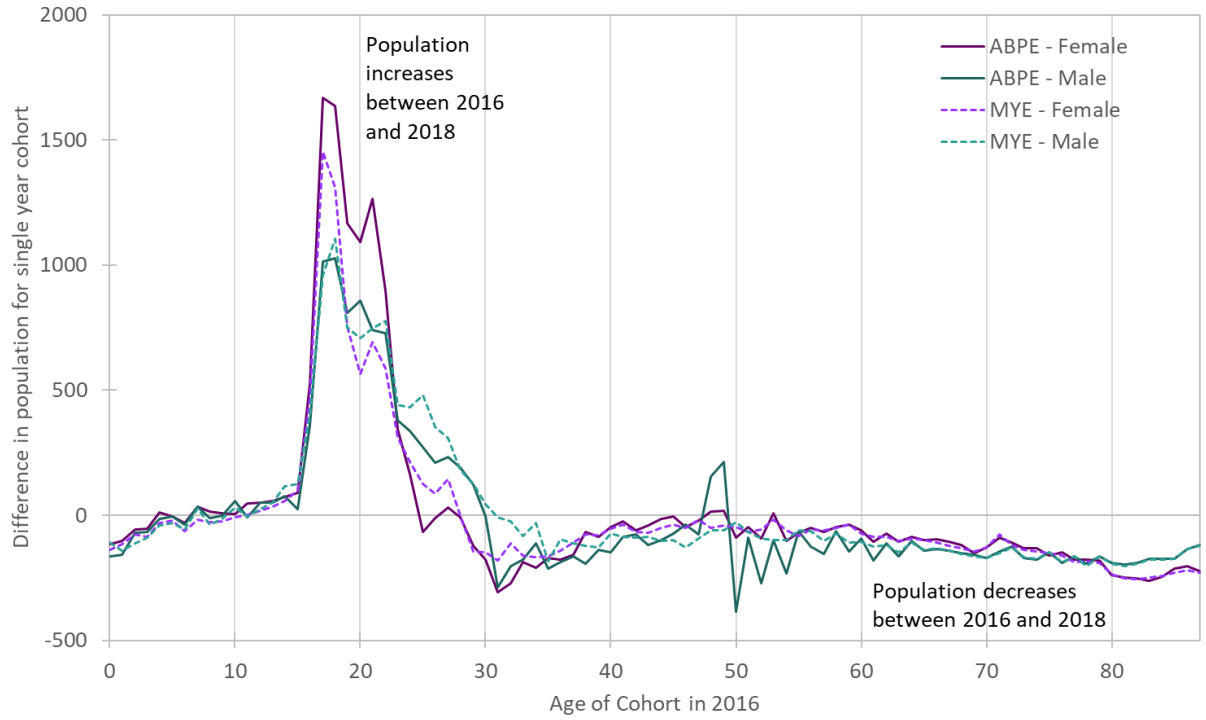
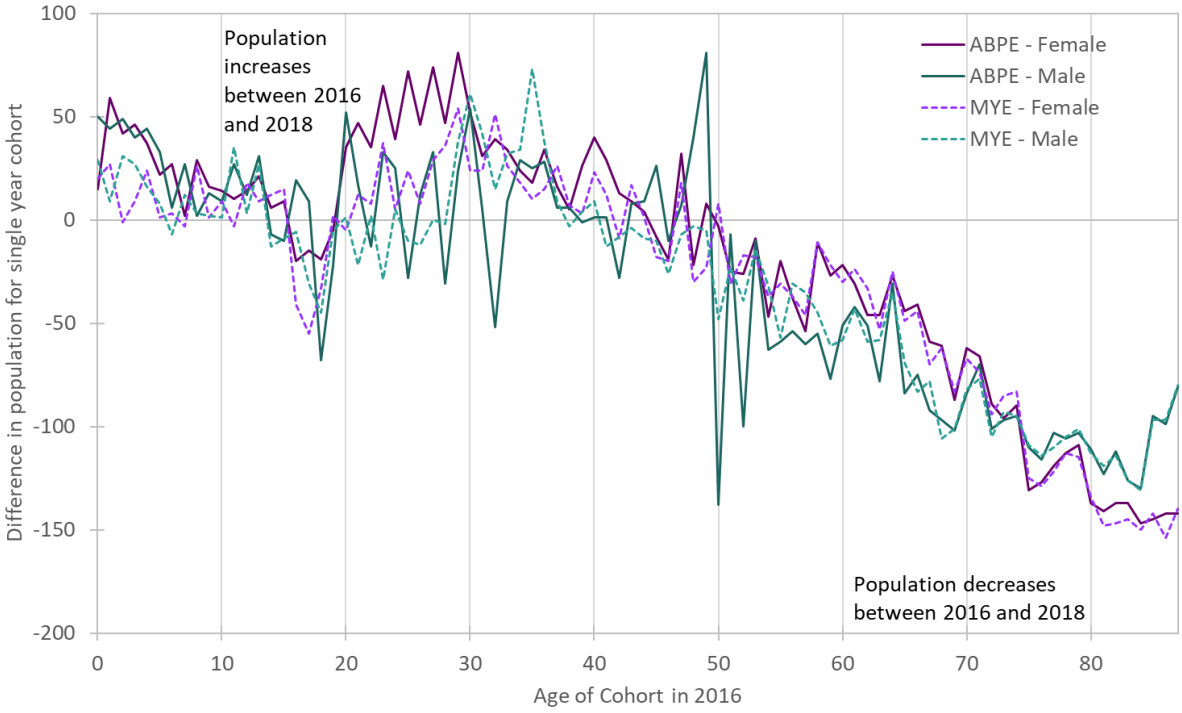


Figure 14 shows a similar comparison for North Lanarkshire. Again there are fluctuations in the MYE, and these are broadly tracked by the ABPE. The large increases around age 20 that were observed for Glasgow City are not seen in North Lanarkshire as North Lanarkshire does not attract a large student population[8]. However, around age 20–30, the ABPE still shows larger increases for females compared with the MYE.

[7] Source: https://www.hesa.ac.uk/data-and-analysis/students/whos-in-he.
[8] Glasgow City has five higher-education institutions and three further-education colleges, while North Lanarkshire has no higher-education institutions and one further-education college (see www.sfc.ac.uk/funding/universities-we-fund.aspx and www.sfc.ac.uk/funding/colleges-we-fund.aspx).

Figure 14: Difference in population changes between 2016 to 2018 by cohort and sex, North Lanarkshire.



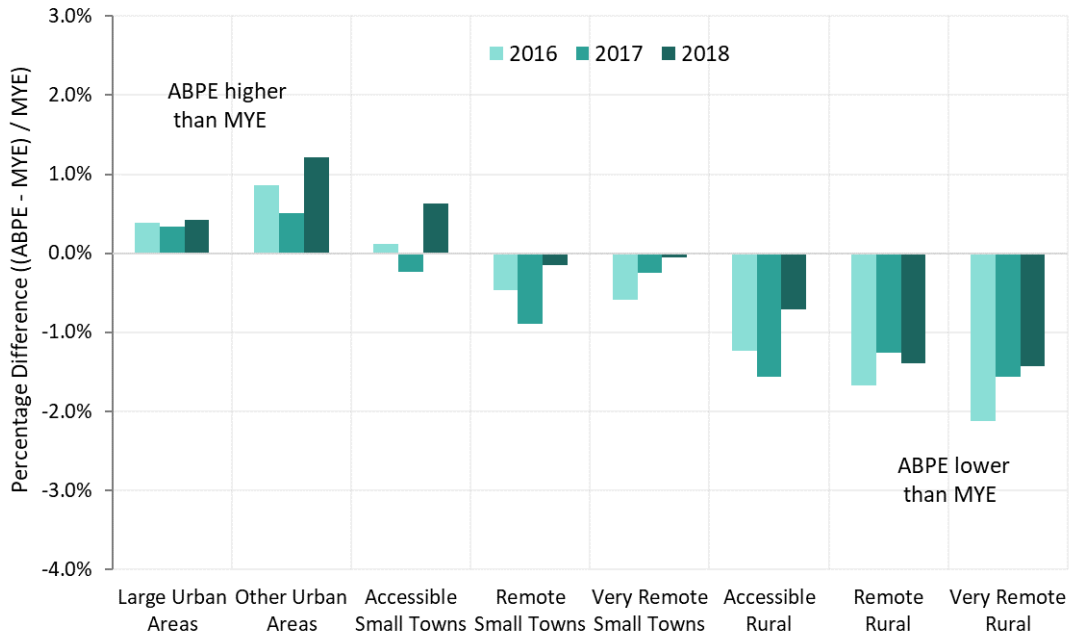
The changes between the ABPE and MYE will still need further investigation for various age ranges for council areas. The statistical research will need to delve deeper in the future but the ABPE in general tends to mirror the population changes shown in the MYE between 2016 and 2018.

A set of interactive charts for council areas have been published as part of the statistical outputs.

Urban-Rural Classification

The comparison between the ABPE and the MYE, broken down by the Scottish Government Urban-Rural Classification (2016) is shown in Figure 15. Note that a change in methodology for mapping people to urban–rural areas, with a switch to using data zones instead of postcode for urban-rural classification, has resolved some of the discrepancies reported in the 2016 ABPE version 1 publication.

Figure 15: Percentage difference between ABPE and MYE by 8-fold urban–rural classification, 2016–2018.



Overall, averaged across the three years, the ABPE was 0.6% higher than the MYE for urban areas, and 0.8% lower than the MYE across other areas. It might be expected that large urban areas have the greatest changes among their population, and hence the largest fluctuation in comparisons between the estimates. However, large urban areas have very similar comparisons across the three years. Other areas tend to have relatively higher comparisons in 2018 than in 2016. This ties in with the finding above that the council areas containing the four largest cities tended to show similar changes in the ABPE and MYE over the three years, while for most other council areas the ABPE tended to increase more than the MYE over this time.

Note that in 2017 Fife accounted for 12.8% of the population in other urban areas, 12.7% of the accessible small towns population and 10.2% of the accessible rural population. The under-coverage in Fife in 2017 may therefore explain why the 2017 figures are lower for the ABPE than other years (relative to the MYE) for these areas.

Figure 16: Percentage difference between the ABPE and MYE by 8-fold urban–rural classification for females.

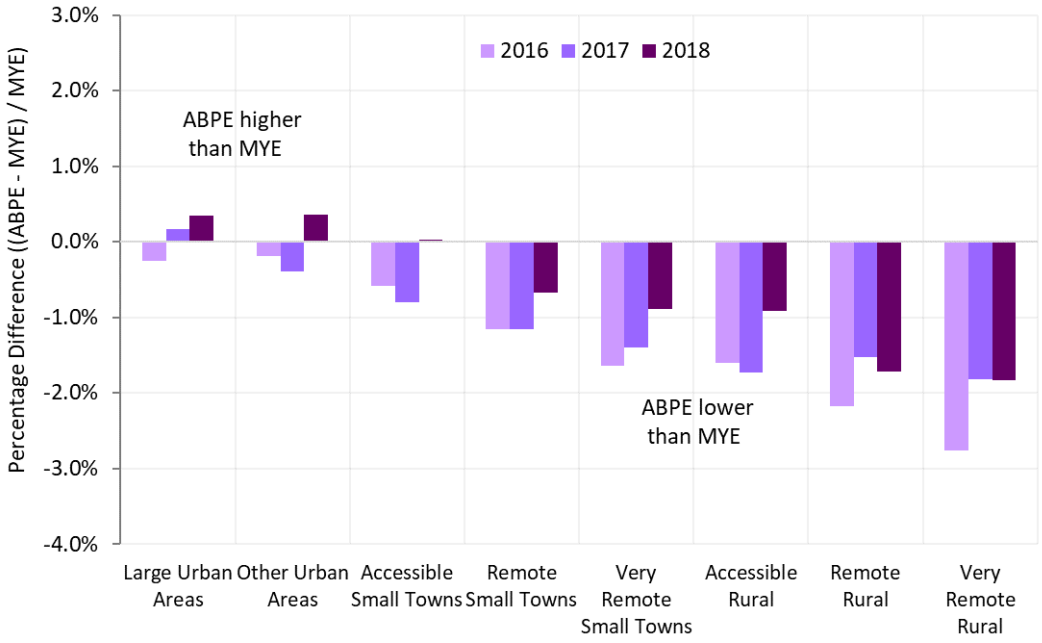
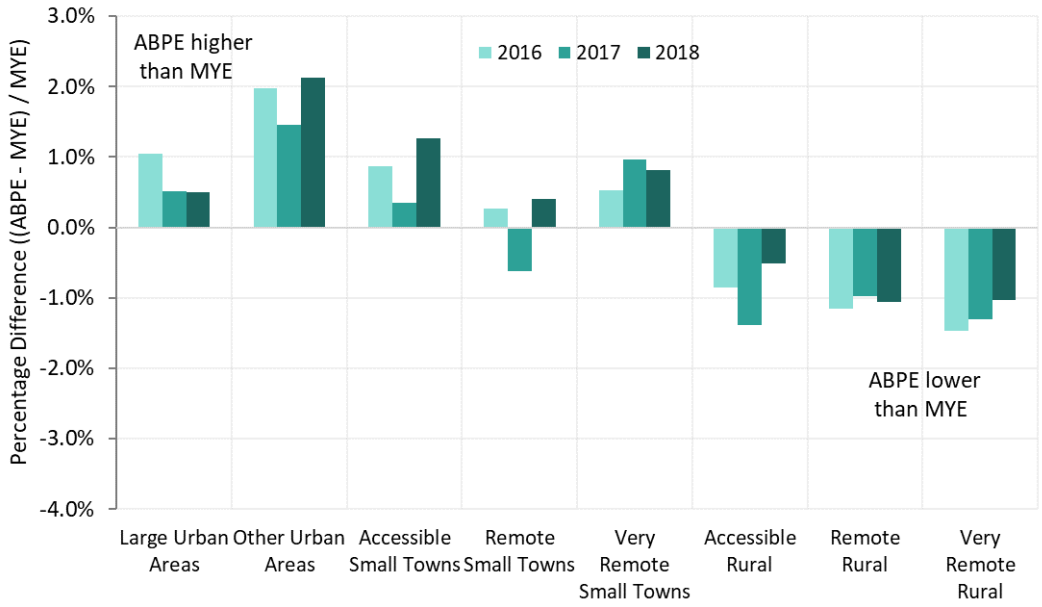


Figure 17: Percentage difference between the ABPE and MYE by 8-fold urban–rural classification for males.



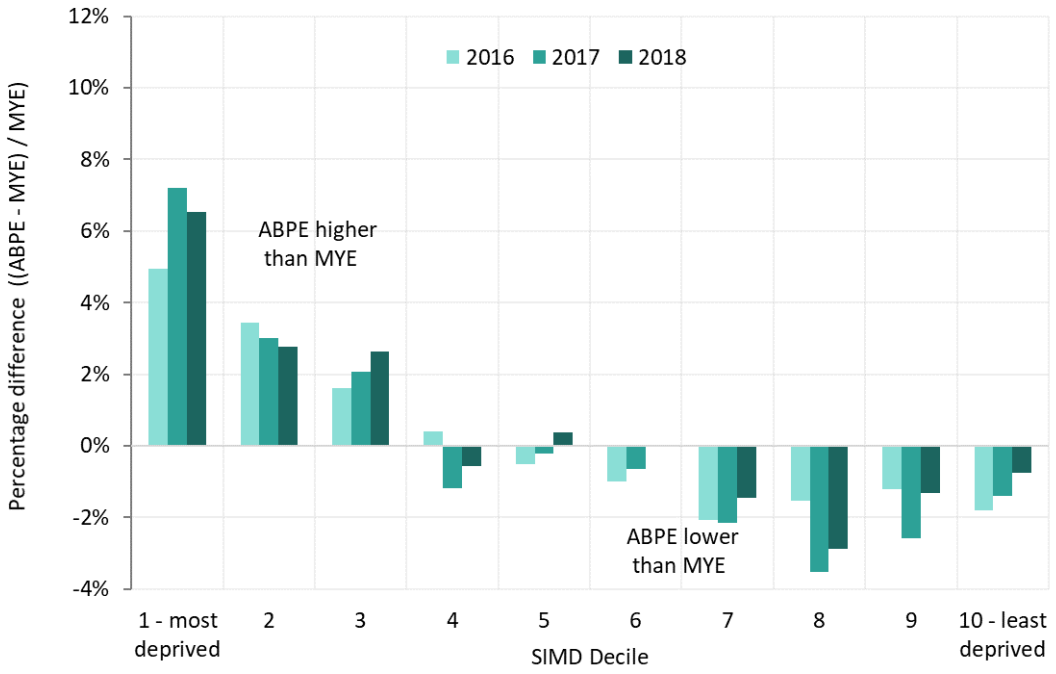
A breakdown by sex in figures 16 and 17 shows similar urban rural patterns for males and females. For males, the ABPE tends to be higher relative to the MYE than for females, reflecting the overall trend discussed in previous sections. For both males and females, increasingly remote and rural areas tend to show the ABPE being increasingly lower than the MYE, although there is more fluctuation among males.

The interactive charts provide further detail, including breakdowns by age. These show that the effect for increasingly remote and rural areas is strongest among the 30–59 year age group. Rural areas tend to have older populations that urban areas. It was seen above that among older people the ABPE tends to be lower than the MYE. Therefore, in rural areas it might be expected that the ABPE is lower than the MYE, simply as a result of this. However, as the pattern also exists within certain age bands (in particular 30–59) it is not the case that the overall pattern is due to differences in the population structure in urban and rural areas.

Scottish Index of Multiple Deprivation (SIMD)

The comparison between the ABPE and the MYE, broken down by 2016 SIMD deciles is shown in Figure 18. The ABPE tends to be higher than the MYE for the most-deprived areas, and lower for less-deprived areas.

Figure 18: Percentage difference between ABPE and MYE by SIMD Decile, 2016–2018.



Figures 19 and 20 show that the effect of SIMD on the comparisons is larger for males than females (note the differing y-axis scales). Broadly similar patterns of percentage difference occur cross the three years.

Figure 19: Percentage difference between ABPE and MYE by SIMD Decile for females, 2016–2018.

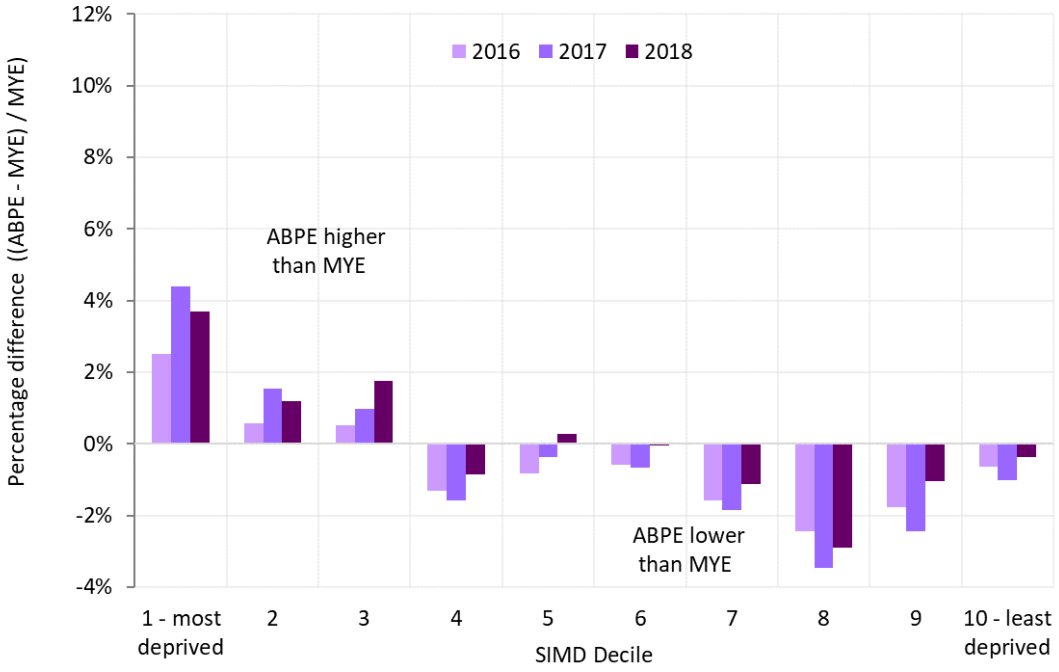


Figure 20: Percentage difference between ABPE and MYE by SIMD Decile for males, 2016–2018.

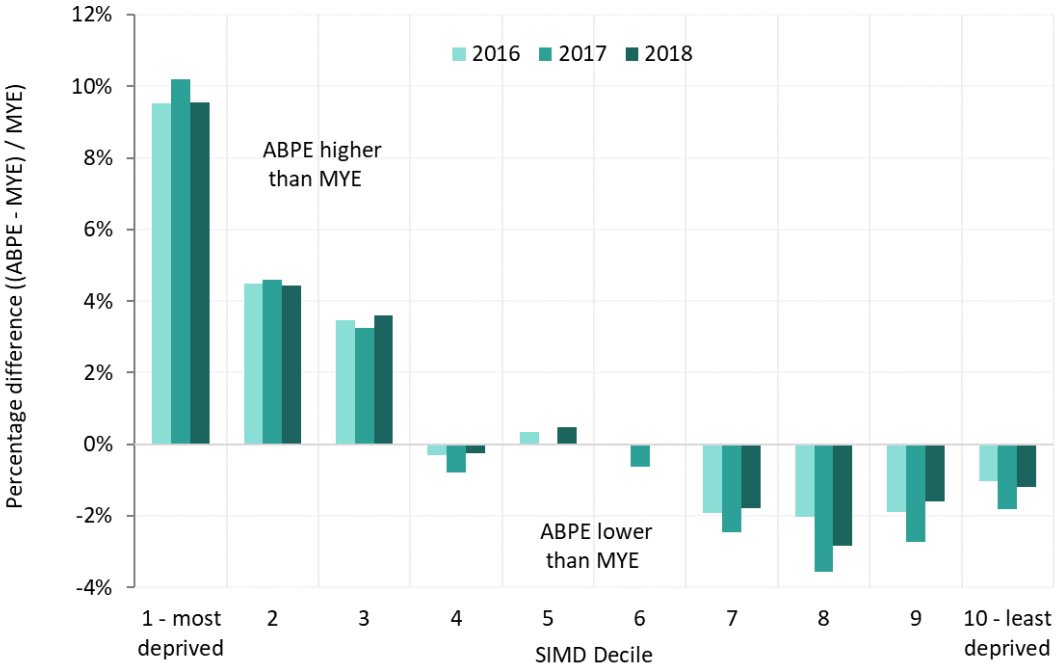
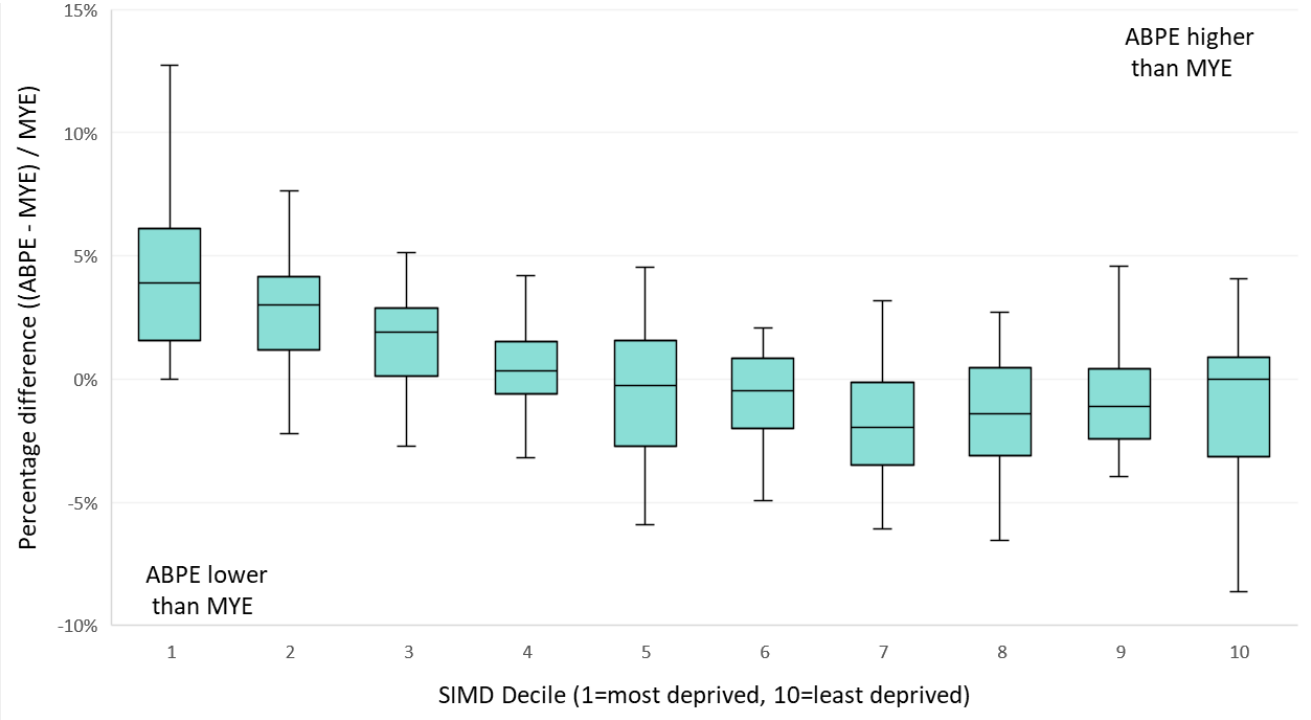


Figure 21, shows that councils tend to reflect the same patterns. In other words, the ABPE being higher than the MYE in the most deprived areas (deciles 1 to 3) is not particularly due to a small number of council areas. The ABPE is higher than the MYE in almost all council areas for SIMD deciles 1 and 2 and in more than 75% of councils for SIMD decile 3 (the boxes are above the x-axis).

Figure 21: Distribution of council area differences between ABPE and MYE by SIMD decile, 2018



Figures 19 and 20 showed that the SIMD effect is stronger for males than for females. To explore this further, figures 22 to 24 show the distribution of council area differences for males of particular ages across the SIMD deciles. This helps show whether the effect is being driven by a particular age group, and/or by a small number of council areas that happen to have a large population in deprived areas. Figure 22 shows that for males aged 0–29 there is a mild effect. For SIMD decile 1, the box is above 0% indicating that more than three quarters of council areas have the ABPE higher than the MYE in the most deprived areas. For SIMD deciles 4 to 6 and 9 to 10 the ABPE tends to be closer to the MYE, and the median (the central line in the boxes) is near to zero.

Figure 22: Distribution of council area differences between ABPE and MYE by SIMD deciles for males aged 0–29, 2018.

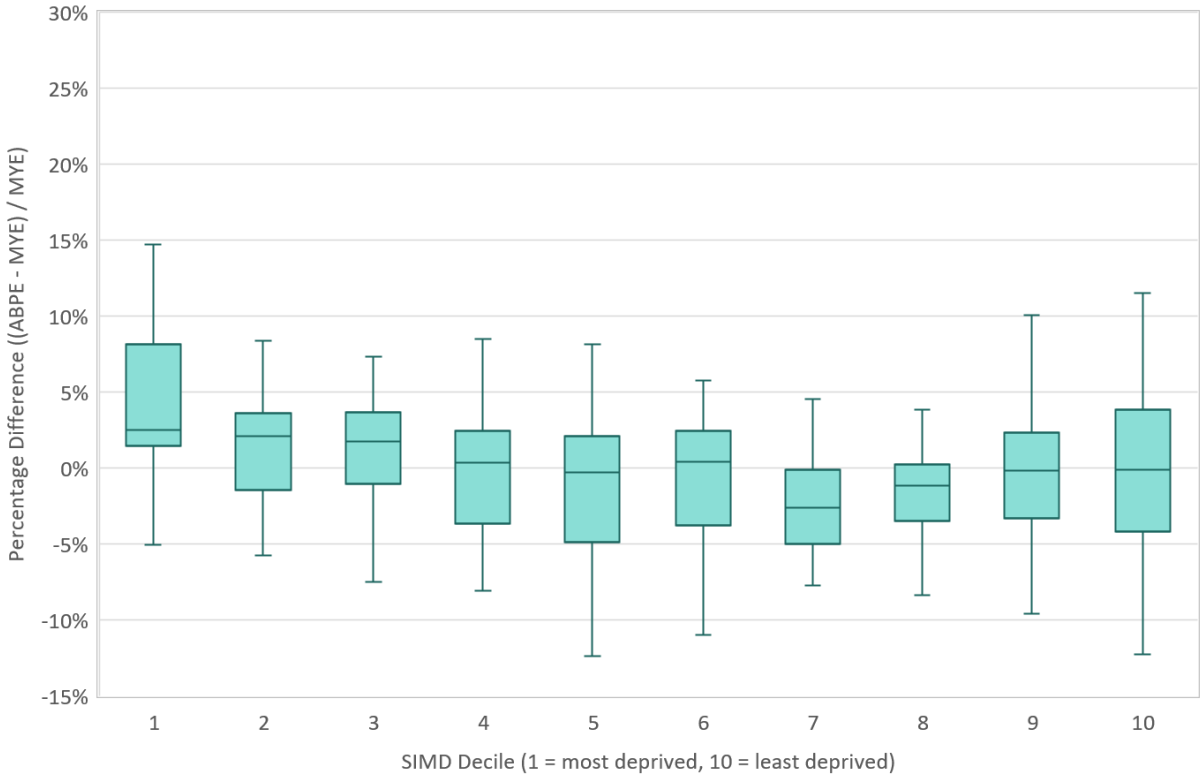


Figure 23 shows the same for males aged 30–59, but for this age group it can be seen that the effect of SIMD is much stronger. For SIMD deciles 1 and 2 all council areas are higher for ABPE than MYE as the minimum percentage difference is above 0%. For SIMD decile 3 there are only two council areas below 0%. Furthermore, the differences tend to be larger than for those aged 0–29. Only SIMD deciles 7 to 9 have the median (the central line in the boxes) showing the ABPE below the MYE, that is, less than 0%. For those aged 0–29 the median had dropped to around 0% by SIMD decile 4.

Figure 23: Distribution of council area differences between ABPE and MYE by SIMD deciles for males aged 30–59, 2018.

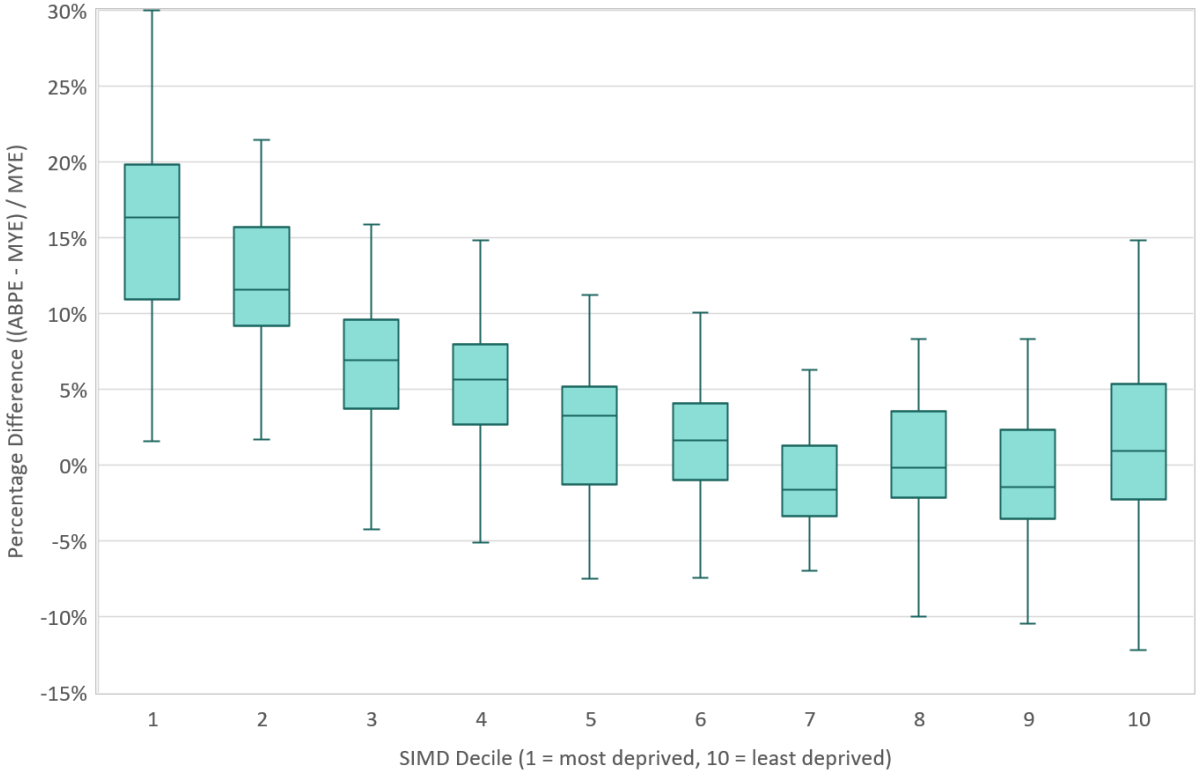
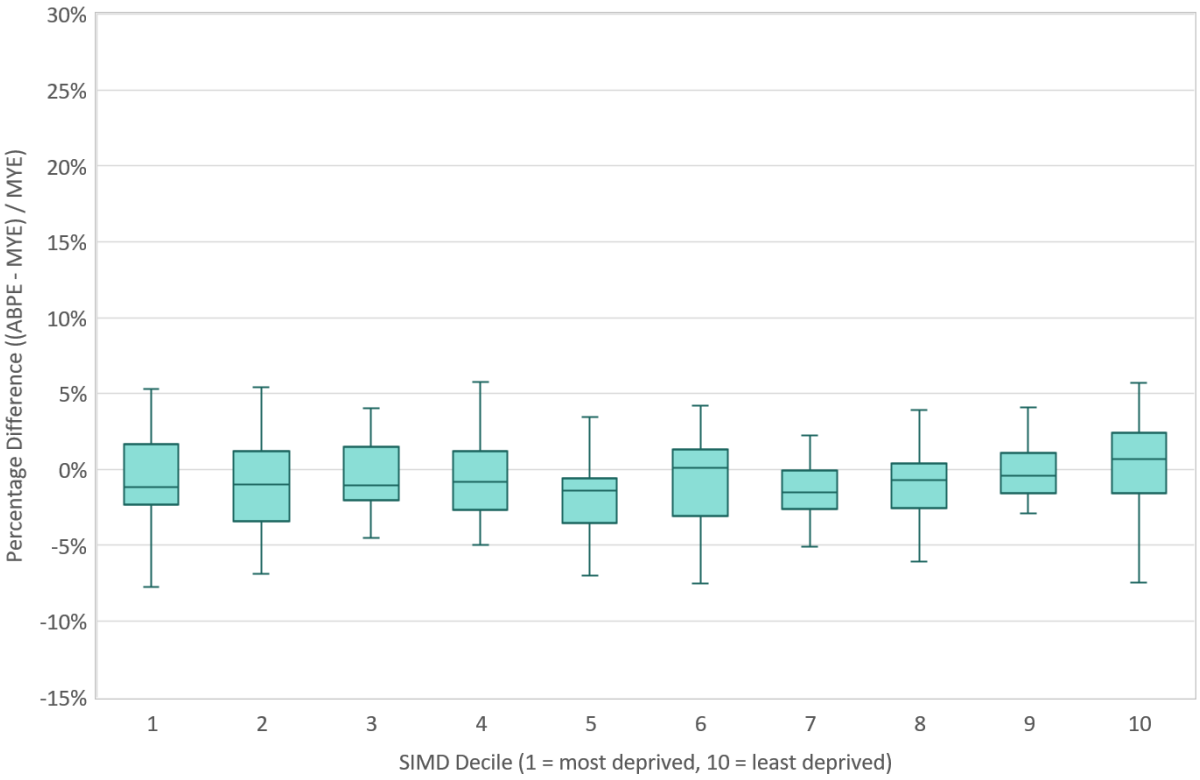


Figure 24: Distribution of council area differences between ABPE and MYE by SIMD deciles for males aged 60+, 2018.



Finally, Figure 24 shows males aged 60 and above. The range here is much smaller than for the other age groups, and the trend is slightly reversed, with SIMD decile 10 being the only decile having the median line above 0%

This shows that the observed SIMD trend (ABPE higher than MYE for deprived areas and lower for less-deprived areas) is mainly due to males aged around 30–59. Furthermore, the SIMD trend is not due to a few council areas, but is present across Scotland. Conversely, this also indicates that the ABPE for males being higher than the MYE around this age range (Figure 7) is mainly due to more males in deprived areas across Scotland appearing on the ABPE.

Limitations of analysis

These outputs are statistical research and the methods will be reviewed for future publications. Although not all people who appear on the datasets used are included in the SIDD, it is likely that there are still people living in Scotland who do not appear on any of the datasets used, as they do not interact with these organisations. Conversely, not everyone who appears on the SIDD will be living in Scotland on the reference date. NRS will be looking into possible new administrative datasets that could help with the coverage of Scotland's population.

The methodology of this project will continue to be refined in the future. It is only when the ABPE is compared with the corresponding census, that we may be able to investigate further some of these differences.

7. Progression on Future Developments/Improvements or Changes in this publication

In our previous publication, we hoped to explore other datasets and make the business rules more specific for different age groups or different geographies. We have been looking into trying to have an activity dataset similar to the one we have for health for benefit data, that is the last interaction not the actual benefit information. This process is at an exploratory stage at the time of this publication. We have made changes to our business rules on age groups and have learned more about geographies relating to some large communal establishments, for example prisons.

We have been sharing the findings of the 2016 statistical research with data users throughout 2021. Presentations have been given to various users; academics, government statistical producers and the main demographic users in Scotland. We are keen to continue this engagement with the research based on three years of data, as we have been able to extend our range of outputs, with more interactive charts.

Currently the estimates are counts of persons who appear on the SIDD. More-sophisticated estimates could make use of the SIDD data, but not be a direct count. A study has started to examine the possibility of using estimation techniques. These techniques include dual system estimation and weighted class estimates. This will require an extensive literature review of the survey landscape of Scotland, and an understanding of UK and international comparisons. This study will be worked on throughout 2022.

8. Future Developments

This statistical research has highlighted some more areas that require further analysis.

- Datazone level information has been used to understand some of our findings in this publication. More work on this geographic area is needed, to assess and understand the quality of the outputs before they can be published.
- We are also keen to start looking at migration between years and hope to have this analysis in the next publication.
- There is also a plan to develop administrative-data based estimates on occupied dwellings. The research for this has begun and will continue into 2022.

9. Background Note

Governance

The governance arrangements for this project were covered in Section 3 in the previous publication. There have been some updates during 2021.

As part of our annual review process an updated Data Protection Impact Assessment (DPIA) has been has been published with this report.

A number of updates to this project have been submitted to the Public Benefit and Privacy Panel for Health and Social Care (PBPP-HSC) and the Statistics Public Benefit and Privacy Panel (SPBPP) covering changes in personnel and for the extension of the data acquisition due to the census moving from 2021 to 2022. The related documentation, such as the data sharing agreements, DPIA, training certificates, and so on, were also reviewed.

The datasets used in this publication are:

- Further Education Statistics (SFC)
- Health Activity (PHS)
- Higher Education Statistics Agency (HESA)
- National Health Service Central Register – NHSCR (NRS)
- Register of Electors (EROs)
- Scottish Pupil Census (SG)
- Vital Events: Birth Registrations (NRS)
- Vital Events: Death Registrations (NRS)
- Vital Events: Marriage Registrations (NRS)
- Vital Events: Civil Partnership Registrations (NRS)

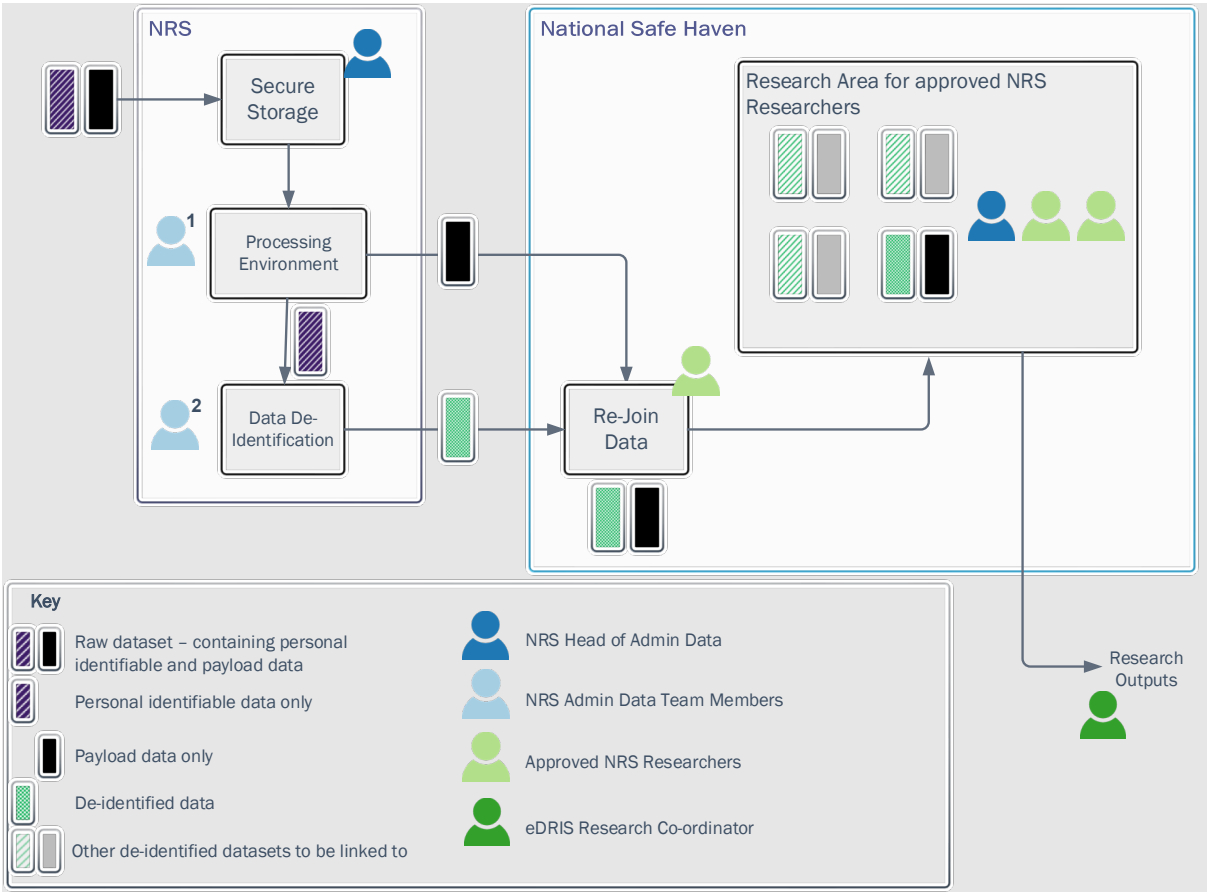
The datasets all relate to slightly different time periods, though covering 2016, 2017 and 2018.

Datasets are sent securely to NRS from data suppliers. They are transferred one at a time to a secure area on the NRS servers to be processed. Once in the secure area the data are quality assured and then de-identified, that is, altered in such a

way that the original values cannot be recovered from the result, thus protecting the identity of the data subjects.

Each dataset is separated into two parts: personal and payload. The personal information contains names, date of birth, postcode and/or address. This goes through the de-identification process. All other information is placed in the payload dataset. While there are two health datasets in this project, they do not include individuals' medical information.

Figure 25: Overview of the flow of data.



The de-identified personal data is then transferred to the National Safe Haven. The datasets are rejoined and analysed by another statistician, who was not involved in their preparation. This method of separating the processes and limiting the staff who see the dataset at different stages increases the security in processing individuals' data. Figure 25 shows this process.

Once the data sources have been processed they are linked as described in the methodology section.

Quality Assurance of Administrative Data (QAAD)

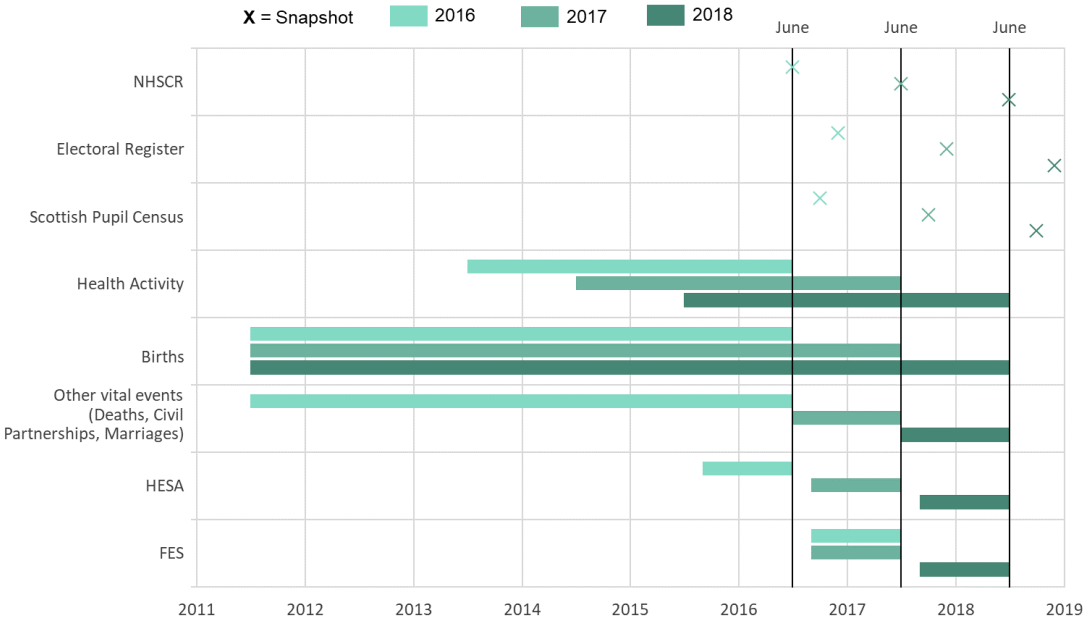
Administrative datasets are collected for a primary purpose for the organisation who owns them. In this project, the data was being used in a different way from the primary purpose. We worked closely with the data suppliers to gain a clear understanding of the strengths and limitations of each dataset in relation to this project. In addition, the completeness and quality of the data was reviewed when it was received from the data suppliers. More detailed information on each dataset has been provided in the QAAD report.

This information helped to support the choice of business rules described in the methodology section

Administrative Datasets Timeline

Several administrative data sources have been used in the production of these population estimates. While the estimates use 30 June as a reference data, the data sources in this publication cover different time periods and have a degree of lag between the time period covered and when they are available. The time periods covered by each dataset are shown in Figure 26.

Figure 26: Time period covered by each dataset.



The timeliness of data is an issue. Although most of the datasets were received before the end of 2016 for ABPE 2016, the datasets covering students (HESA and FES) were not available until spring 2017. The de-identification process takes six months, limiting publication to one year after the reference date at the earliest. Considering the administrative data based population estimates for 2022 as an example, it would be approximately autumn/winter 2023 before all the datasets were available to analyse. NRS will work with our data suppliers to explore ways to improve this.

There are two QAAD reports covering this statistical research publication. The data used in this research for 2016 is the same as that used for the previous publication. Whilst the underlying data has not changed, the business rules used to create the population estimates have changed.

- Quality Assurance of Administrative Datasets, 2016
- Quality Assurance of Administrative Datasets, 2017 and 2018

Stakeholder Engagements

Stakeholder engagement has always been important to this project; the NRS Admin Data team in 2017 embarked on a wide ranging process of stakeholder engagement. The details of this are published on our website.

The Population And Migration Statistics Committee (PAMS) is one of the main demographic users' groups in Scotland. Any substantial progress was supplied to this group.

We have presented the statistical research of ABPE 2016 to users throughout 2021. The feedback will be used to shape future statistical research in this area.

Following this publication, we wish to discuss the findings of this research with as many users as possible. If you have any comments or would like to be informed of future stakeholder events, then please register your interest under the topic title: demography at http://www.gov.scot/scotstat.

Revisions

This statistical research is expected to be revised in subsequent publications due to improvements in methodology. Those subsequent publications will supersede this publication. We will aim to update the results for previous years with the new methodology to allow trend analysis.

Methodology version 1 – Administrative Data Based Population Estimates Scotland 2016

Methodology version 2 – Administrative Data Based Population Estimates v2 Scotland 2016–2018

Revisions and corrections to previously published statistics are dealt with in accordance with the Scottish Government Statistician Group corporate policy statement on revisions and corrections.

Comparison of administrative data based population estimates within the UK

Similar approaches to producing population estimates using administrative data are being conducted by other statistical agencies nationally and internationally. Though the approaches are similar, the major differences are the different types of administrative data that are available to each organisation.

In the UK, the Office for National Statistics (ONS) have a programme of work called the Transformation of Population and Migration Statistics. This work is aiming to put administrative data at the core of what they do.

ONS have release published Developing admin-based population estimates, England and Wales: 2016 to 2020. They currently have two methods for producing administrative data based population estimates. They report on progress toward understanding the stability and behaviour of these.

The Northern Ireland Statistics and Research Agency (NISRA) produced population estimates based on data gathered from statistical censuses and surveys, and data extracted from its own and other organisations' administrative or management systems. NISRA reviewed 'The use of administrative data in population estimates' and published their initial findings for Northern Ireland in 2014.

Comparison of administrative data based population estimates outwith the UK

Internationally, New Zealand's Stats NZ have developed their own research environment, the Integrated Data Infrastructure and have used it to produce various administrative data based population estimates.

The Australian Bureau of Statistics (ABS) have been comparing counts from administrative data with official ABS population counts for 2016. This work is to inform how administrative data can support future Australian censuses.

Statistics Canada have also been looking at creating a population spine, initially using data from 2011 administrative data sources. They also use administrative data sources in the production of official statistics including population and demography.

Ireland's Central Statistics Office (CSO) have produced migration estimates for Ireland from administrative data sources in April 2021.

The NRS National Statistics publications used for benchmarking purposes are:

- Mid-Year Population Estimates (MYE)
- Centenarians in Scotland, 2010 to 2020

These are the sources that should be used when doing any research or analysis using population statistics, and the outputs presented here **should not** be used as an alternative source.

Supporting Documentation

This publication has a number of supporting documents, these can be found on our website:

- Administrative Data Based Population Estimates v2 Scotland 2016–2018: Methodology Report
- Administrative Data Based Population Estimates, Scotland 2017 and 2018 – Quality Assurance of Administrative Dataset
- Data Protection Impact Assessment (DPIA) – Administrative Data Based Population and Household Estimates Project
- Voluntary Adopter of Code of Official Statistics Statement
- Administrative Data Based Population Estimates, Scotland 2016 to 2018 – Tables
- Administrative Data Based Population Estimates, Scotland 2016 to 2018 – Charts
- Administrative Data Based Population Estimates, Scotland 2016 to 2018 – Interactive Charts

10. Glossary

The table below provides a description of the abbreviations used in this document.

Abbreviation	Description
ABPE	Administrative Data Based Population Estimates
CSO	Central Statistics Office [Ireland]
DPIA	Data-Protection Impact Assessment
ERO	Electoral Register Officer
FES	Further Education Statistics
GDPR	General Data Protection Regulation
HESA	Higher Education Statistics Agency
MYE	Mid-Year Estimate
NHSCR	National Health Service Central Register
NISRA	Northern Ireland Statistics and Research Agency
NRS	National Records of Scotland
ONS	Office for National Statistics [UK]
PAMS	Population and Migration Statistics Committee
PBPP-HSC	Public Benefit and Privacy Panel for Health and Social Care
PHS	Public Health Scotland
QAAD	Quality Assessment of Administrative Data
SFC	Scottish Funding Council
SG	Scottish Government
SIDD	Scotland's Integrated Demographic Dataset
SIMD	Scottish Index of Multiple Deprivation
SPBPP	Statistics Public Benefit and Privacy Panel

11. Notes on statistical publications

Statistical Research

This publication presents statistical research and the methodology is still under development. We welcome any feedback from users on ways in which the methodology or data sources may be developed to improve the quality of these statistics in future years.

Information on background and source data

Further supporting documentation is published alongside this publication on the NRS website. The hyperlinks link can be found at the end of Section 9

National Records of Scotland

We, the National Records of Scotland, are a non-ministerial department of the devolved Scottish Administration. Our aim is to provide relevant and reliable information, analysis and advice that meets the needs of government, business and the people of Scotland. We do this as follows:

Preserving the past – We look after Scotland's national archives so that they are available for current and future generations, and we make available important information for family history.

Recording the present – At our network of local offices, we register births, marriages, civil partnerships, deaths, divorces and adoptions in Scotland.

Informing the future – We are responsible for the Census of Population in Scotland which we use, with other sources of information, to produce statistics on the population and households.

You can get other detailed statistics that we have produced from the Statistics section of our website. Scottish Census statistics are available on the Scotland's Census website.

We also provide information about future publications on our website. If you would like us to tell you about future statistical publications, you can register your interest on the Scottish Government ScotStat website.

You can also follow us on twitter @NatRecordsScot

Enquiries and suggestions

Please get in touch if you need any further information, or have any suggestions for improvement.

Lead Statistician: Lindsay Bennison

Statistics Customer Services telephone: (0131) 314 4299

E-mail: statisticscustomerservices@nrscotland.gov.uk

For media enquiries, please contact: scotlandscensus@nrscotland.gov.uk